

Présentation à partir d'exemples élémentaires de la problématique des plans d'expériences

A. Kobilinsky

le 17 Mars 1992

Labo de Biométrie.
INRA. Route de St-Cyr.
F78026 VERSAILLES Cedex.

Table des matières

1	Sondage pour l'estimation de la qualité d'un stock de viande	3
1.1	Détermination du plan de sondage optimal	3
1.2	Randomisation et modèle associé	10
1.3	Estimation des paramètres v_A, v_B, v_E	15
2	Etude d'un facteur à 2 niveaux	17
2.1	Exemple et modèle	17
2.2	Rappel sur l'estimation des paramètres	18
2.3	Randomisation ¹	20
2.4	Plan optimal	25
2.5	Efficacité et estimabilité	26
3	Etude de 2 facteurs à 2 niveaux	27
3.1	Exemples	27
3.2	Modèle et effets factoriels	29
3.3	Estimation et variance d'estimation	31
3.4	Plan optimal	32
3.5	Variance dans les dispositifs déséquilibrés.	34
3.6	Estimation dans le cadre du modèle additif	36
3.7	Plans optimaux pour la minimisation du biais	44
4	Régression simple	44
4.1	Modèle et estimation	44
4.2	Plan optimal	46
4.2.1	Plan optimal pour l'estimation de la pente	46

1. Ce paragraphe plus technique peut être laissé de côté dans une première lecture

4.2.2	Plan optimal pour l'estimation de la réponse théorique	47
4.2.2.1	Critère G et théorème de Kiefer-Wolfowitz	47
4.2.2.2	Critère de variance moyenne	48
4.3	Prise en compte du biais	49

Liste des tableaux

1	Ecart à l'optimalité du plan à effectifs égaux dans le cas de variances inégales	26
2	Efficacité dans le plan pour 1 facteur	27
3	Traitements dans des plans pour 2 facteurs à 2 niveaux	28
4	Le plan d'expérience	28
5	Matrice H de passage des effets factoriels aux effets traitements	30
6	Variances des effets dans le cas non standard	33
7	Efficacité avec les effectifs n_1, n_1, n_2, n_2 , où $n_1 = pn/4$	34
8	Efficacité avec trois des effectifs égaux et le quatrième égal à $pn/4$	35
9	Un plan fortement déséquilibré	40
10	Comparaison des estimateurs des modèles avec et sans interaction	43

Les développements qui suivent, sur l'estimation de la quantité de lipide dans de la viande, puis sur les plans pour 1 ou 2 facteurs à deux niveaux et enfin sur la régression simple linéaire et quadratique, ont pour objectif d'introduire les notions de base utilisées dans la planification d'expériences.

Un bon plan de sondage ou un bon plan d'expérience est celui qui, pour des ressources finies, permet d'obtenir le maximum d'information. Si les objectifs à atteindre sont fixés à priori, le bon plan est alors celui qui utilise un minimum de ressources. L'exemple qui suit illustre ce propos dans le cas simple où l'objectif poursuivi est l'estimation d'un unique paramètre.

1 Sondage pour l'estimation de la qualité d'un stock de viande

1.1 Détermination du plan de sondage optimal

Dans un contrôle de réception de cartons de viande, on désire estimer le pourcentage de graisse moyen. A cet effet, on opère des prélèvements dans n_A cartons choisis au hasard, à raison de n_B/n_A prélèvements par carton. Les n_B prélèvements ainsi effectués sont broyés ensemble et on effectue ensuite n_E mesures distinctes du pourcentage de graisse dans le broyat. Il est naturel de se demander comment on doit choisir n_A, n_B et n_E pour que l'estimation soit la plus précise possible, étant entendu qu'on se fixe un coût global maximum c_0 .

On note c_B le coût d'un prélèvement
 c_A le coût additionnel d'ouverture d'un carton
 c_E le coût d'une mesure

La contrainte du coût est donc

$$c_A n_A + c_B n_B + c_E n_E \leq c_0 \quad (1)$$

On suppose que le pourcentage de graisse y_{ij} pour le $j^{\text{ème}}$ prélèvement du carton i suit le modèle²

$$y_{ij} = \mu + A_i + B_{ij}, \quad \text{où} \quad (2)$$

2. Nous verrons plus loin qu'un modèle similaire peut être obtenu en s'appuyant sur le seul fait qu'on a tiré au hasard les cartons, puis les prélèvements dans les cartons, les tirages étant effectués avec probabilités égales et sans remise (tirage PESR)

- μ est le pourcentage de graisse moyen que l'on veut estimer
- A_i l'effet du carton i
- B_{ij} l'effet du prélèvement j dans le carton i .
- les effets A_i et B_{ij} sont centrés et non corrélés entre eux $\left(E(A_i) = 0, E(B_{ij}) = 0, \text{cov}(A_i, A_{i'}) = 0 \text{ pour } i \neq i', \text{etc} \dots \right)$
- les A_i ont une même variance v_A et les B_{ij} une même variance v_B :

$$\begin{aligned} \text{var}(A_i) &= v_A \\ \text{var}(B_{ij}) &= v_B \end{aligned} \quad (3)$$

Dans ces conditions, le pourcentage de graisse moyen dans le broyat est

$$y_{..} = \mu + A_{.} + B_{..}$$

expression dans laquelle les points indiquent que l'on a effectué une moyenne sur l'indice correspondant :

$$y_{..} = \frac{1}{n_B} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b/n_a} y_{ij}.$$

Le résultat Y_k de la $k^{\text{ème}}$ mesure s'écrit

$$Y_k = y_{..} + E_k \quad (4)$$

où l'erreur de mesure E_k est centrée et a une variance v_E :

$$\begin{aligned} E(E_k) &= 0 \\ \text{var}(E_k) &= v_E \end{aligned} \quad (5)$$

On estime le pourcentage de graisse μ par la moyenne de ces mesures Y_k :

$$\hat{\mu} = Y_{.} = y_{..} + E_{.} \quad (6)$$

soit

$$\hat{\mu} = \mu + A_{.} + B_{..} + E_{.}$$

On a $E(\hat{\mu}) = \mu$, donc $\hat{\mu}$ est une estimation sans biais de μ . L'imprécision sur $\hat{\mu}$ est donnée par la variance de $\hat{\mu}$. Compte tenu de la nullité des covariances entre les effets A_i , B_{ij} et les erreurs E_k , cette variance s'écrit :

$$\text{var}(\hat{\mu}) = \text{var}(A_{.}) + \text{var}(B_{..}) + \text{var}(E_{.})$$

soit

$$\text{var}(\hat{\mu}) = \frac{v_A}{n_A} + \frac{v_B}{n_B} + \frac{v_E}{n_E} \quad (7)$$

Pour avoir l'estimation la plus précise possible à coût c_0 fixé, on doit donc chercher les valeurs de n_A , n_B , n_E qui minimisent cette variance tout en respectant la contrainte de coût (1). Il y a évidemment d'autres contraintes à respecter puisque n_A , n_B/n_A et n_E doivent être des entiers supérieur à 1, que n_A ne peut dépasser le nombre total N_A de cartons et de même que n_B/n_A ne peut dépasser le nombre N_B/N_A maximum de prélèvements possibles dans un carton.

On fera dans un premier temps abstraction du fait que n_A , n_B/n_A et n_E doivent être des entiers et on ignorera les contraintes qui fixent les limites supérieures pour n_A et n_B/n_A . Pour minimiser $\text{var}(\hat{\mu})$, on utilise alors la méthode des multiplicateurs de Lagrange (cf par exemple Gill, Murray, Wright 1981 *Practical optimization*. Academic Press, ainsi que Pchénitchny, Danilin 1977 *Méthodes numériques dans les problèmes d'extrénum*, Edition MIR, Moscou) Les contraintes prises en compte

$$\begin{aligned} c_A n_A + c_B n_B + c_E n_E &\leq c_0 \\ n_A &\geq 1, \quad n_E \geq 1 \quad n_B \geq n_A \end{aligned} \quad (8)$$

sont réexprimées sous la forme (standard):

$$\begin{array}{rcll} -c_A n_A & -c_B n_B & -c_E n_E & \geq -c_0 \\ -n_A & +n_B & & \geq 0 \\ n_A & & & \geq 1 \\ & & n_E & \geq 1 \end{array}$$

ou sous forme matricielle:

$$Ax \geq b$$

avec

$$A = \begin{bmatrix} -c_A & -c_B & c_E \\ -1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad x = \begin{bmatrix} n_A \\ n_B \\ n_E \end{bmatrix} \quad b = \begin{bmatrix} -c_0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

Le vecteur colonne $g(x)$ des dérivées partielles de $\frac{v_A}{n_A} + \frac{v_B}{n_B} + \frac{v_E}{n_E}$, par rapport à n_A, n_B, n_E est :

$$g(x) = \begin{bmatrix} -\frac{v_A}{n_A^2} \\ -\frac{v_B}{n_B^2} \\ -\frac{v_E}{n_E^2} \end{bmatrix}$$

Au minimum, ce vecteur doit être une combinaison linéaire à coefficients positifs ou nuls des lignes de A :

$$\begin{bmatrix} -\frac{v_A}{n_A^2} \\ -\frac{v_B}{n_B^2} \\ -\frac{v_E}{n_E^2} \end{bmatrix} = \lambda \begin{bmatrix} -c_A \\ -c_B \\ -c_E \end{bmatrix} + \lambda_B \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + \lambda_A \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \lambda_E \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (9)$$

De plus les coefficients associés aux contraintes non actives, c'est à dire à des inégalités strictes, doivent être nuls, i.e.:

$$\begin{aligned} \lambda &= 0 & \text{si} & \quad c_A n_A + c_B n_B + c_E n_E < c_0 \\ \lambda_B &= 0 & \text{si} & \quad n_B > n_A \\ \lambda_A &= 0 & \text{si} & \quad n_A > 1 \\ \lambda_E &= 0 & \text{si} & \quad n_E > 1 \end{aligned} \quad (10)$$

Dans cas le présent la convexité stricte de la fonction à minimiser sur le domaine convexe défini par les inégalités (8) permet d'affirmer que (9) et (10) sont des conditions suffisantes pour avoir un minimum de la fonction.

Pour chercher le minimum, nous allons d'abord considérer les triplets (n_A, n_B, n_E) pour lesquels seule la première contrainte est active et qui vérifie donc :

$$\begin{aligned} c_A n_A + c_B n_B + c_E n_E &= c_0 \\ n_B > n_A \quad n_A > 1 \quad n_E > 1 \end{aligned} \quad (11)$$

Pour ces triplets, on doit avoir $\lambda_B = \lambda_A = \lambda_E = 0$ et (9) devient:

$$\begin{bmatrix} -\frac{v_A}{n_A^2} \\ -\frac{v_B}{n_B^2} \\ -\frac{v_E}{n_E^2} \end{bmatrix} = \lambda \begin{bmatrix} -c_A \\ -c_B \\ -c_E \end{bmatrix} \iff \begin{cases} n_A^2 = \frac{1}{\lambda} \frac{v_A}{c_A} \\ n_B^2 = \frac{1}{\lambda} \frac{v_B}{c_B} \\ n_E^2 = \frac{1}{\lambda} \frac{v_E}{c_E} \end{cases} \quad (12)$$

En d'autre terme, les carrés n_A^2 , n_B^2 , n_E^2 des nombres de répétitions doivent être proportionnels aux ratios v_A/c_A , v_B/c_B , v_E/c_E des variances par les coûts associés.

Ce résultat n'est pas surprenant. Il indique qu'on a intérêt à multiplier les répétitions pour les sources les plus importantes de variabilité, si le coût associé à chacune de ces répétitions n'est pas trop important.

Pour un triplet (n_A, n_B, n_E) qui vérifie (12), posons $r_B = n_B/n_A$, $r_E = n_E/n_A$. On a donc:

$$r_B = \frac{n_B}{n_A} = \sqrt{\frac{v_B/c_B}{v_A/c_A}} \quad r_E = \frac{n_E}{n_A} = \sqrt{\frac{v_E/c_E}{v_A/c_A}} \quad (13)$$

Pour que les contraintes (11) soient vérifiées, on doit avoir:

$$\begin{aligned} n_A c_A + r_B n_A c_B + r_E n_A c_E &= c_0 \\ r_B > 1 \quad n_A > 1 \quad r_E n_A > 1 \end{aligned} \quad (14)$$

La contrainte $r_B > 1$ est satisfaite si $v_B/c_B > v_A/c_A$, ce qui se produit si l'effet carton est faible, mais le surcoût c_A d'ouverture d'un carton non négligeable. Si les trois autres contraintes sont également satisfaites, on a:

$$\frac{c_0}{c_A + r_B c_B + r_E c_E} > \max \left(1, \frac{1}{r_E} \right) \quad (15)$$

Réciproquement, si $v_B/c_B > v_A/c_A$ et si l'inégalité (15) est vérifiée, l'effectif $n_A = c_0 / (c_A + r_B c_B + r_E c_E)$ vérifie les contraintes (14) et le triplet

$$n_A = \frac{c_0}{c_A + r_B c_B + r_E c_E}, \quad n_B = r_B n_A \quad n_E = r_E n_A \quad (16)$$

qui s'en déduit satisfait aux conditions (8) et (9) requises pour un minimum.

Pratiquement, la condition (15) est vérifiée si d'une part les ratios v_A/c_A , v_B/c_B , v_E/c_E ne sont pas trop différents, d'autre part le coût maximum autorisé c_0 est assez conséquent par rapport aux coûts d'échantillonnage unitaires.

Considérons maintenant les triplets (n_A, n_B, n_E) tels que

$$n_B = n_A \quad n_A > 1 \quad n_E > 1$$

Pour ces triplets, on a $\lambda_A = \lambda_E = 0$ et le système (9) devient

$$\begin{aligned} -\frac{v_A}{n_A^2} &= -\lambda c_A - \lambda_B \\ -\frac{v_B}{n_A^2} &= -\lambda c_B + \lambda_B \\ -\frac{v_E}{n_E^2} &= -\lambda c_E \end{aligned}$$

ou encore

$$\begin{aligned} \frac{v_A/c_A - v_B/c_B}{n_B^2} &= \lambda_B \left(\frac{1}{c_A} + \frac{1}{c_B} \right) \\ n_A^2 &= \frac{1}{\lambda} \frac{v_A + v_B}{c_A + c_B} \\ n_E^2 &= \frac{1}{\lambda} \frac{v_E}{c_E} \end{aligned}$$

La première égalité est satisfaite avec un coefficient $\lambda_B \geq 0$ si

$$v_A/c_A \geq v_B/c_B$$

Les deux autres égalités sont satisfaites avec un $\lambda \geq 0$ si

$$r_E = \frac{n_E}{n_A} = \frac{\sqrt{v_E/c_E}}{\sqrt{\frac{v_A + v_B}{c_A + c_B}}}.$$

Les autres contraintes deviennent alors :

$$n_A(c_A + c_B) + r_E n_A c_E = c_0$$

$$n_A > 1 \quad r_E n_A > 1$$

Elles sont satisfaites pour $n_A = \frac{c_0}{c_A + c_B + r_E c_E}$ si et seulement si:

$$\frac{c_0}{c_A + c_B + r_E c_E} > \max\left(1, \frac{1}{r_E}\right)$$

Minimisation du coût à variance constante

Ce qui précède permet d'obtenir, pour chaque coût plafond c_0 fixé à l'avance, les nombres n_A, n_B, n_E et la variance minimum correspondante pour $\hat{\mu}$ donnée par (7), qui est notée $v_m(c_0)$. Bien entendu la fonction $c_0 \mapsto v_m(c_0)$ est décroissante, puisque toute solution acceptable avec un coût maximum c_0 est a fortiori acceptable avec un coût maximum c'_0 plus grand que c_0 . Il est clair aussi que $v_m(c_0)$ est une fonction strictement décroissante de c_0 , puisque avec un supplément de coût on peut toujours augmenter le nombre n_E de mesure et faire décroître ainsi $\text{var}(\hat{\mu})$ (en faisant abstraction du fait que n_E est entier). Ceci montre d'ailleurs que la contrainte de coût est toujours active au minimum de la variance.

Le coût minimum permettant d'obtenir une variance inférieure ou égale à v_0 est alors la solution de $v_m(c_0) = v_0$. Car si on pouvait trouver une solution donnant une variance inférieure ou égale à v_0 avec un coût c'_0 inférieur, on aurait : $c'_0 < c_0$ et $v_m(c'_0) \leq v_m(c_0)$, ce qui contredirait le fait que v_m est strictement décroissant.

A titre d'exemple, plaçons nous dans le cas décrit par (11) où la contrainte de coût est la seule active. Les valeurs de n_A, n_B, n_E permettant d'avoir une variance minimum sont données par (12). Si on remplace n_A, n_B, n_E par ces valeurs dans les expressions et donnant le coût (11, ligne 1) et la variance (7), on obtient :

$$\frac{1}{\sqrt{\lambda}} (\sqrt{v_A c_A} + \sqrt{v_B c_B} + \sqrt{v_E c_E}) = c_0$$

$$\sqrt{\lambda} (\sqrt{v_A c_A} + \sqrt{v_B c_B} + \sqrt{v_E c_E}) = \text{var}(\hat{\mu}) = v_m(c_0)$$

et donc

$$v_m(c_0) = \frac{(\sqrt{v_A c_A} + \sqrt{v_B c_B} + \sqrt{v_E c_E})^2}{c_0}$$

Le coût minimal pour obtenir une variance v_0 est donc

$$c_0 = \frac{(\sqrt{v_A c_A} + \sqrt{v_B c_B} + \sqrt{v_E c_E})^2}{v_0}$$

Contraintes de nombres entiers

Ce qui précède à surtout pour but de donner les ordres de grandeur respectifs des nombres n_A , n_B , n_E . Une fois trouvées les valeurs réelles (dans \mathbb{R}) minimisant la variance, on prend pour n_A , n_B , n_E des entiers voisins. Si l'on veut être tout à fait rigoureux, on calcule la variance pour tous les triplets d'entiers satisfaisant les contraintes, et on choisit le triplet qui conduit au minimum.

Autres considérations pratiques

En pratique, il est souvent difficile d'effectuer de vrais tirages au hasard. Ainsi, si les cartons sont disposés en palettes, seuls les cartons situés en haut de palette peuvent être sélectionnés. De même l'échantillonnage dans un carton peut être déterminé par un appareil qui effectue des carottes à des endroits très systématiques.

Pour le choix des cartons, il faut randomiser au sein des cartons facilement accessibles, et espérer que ces derniers sont effectivement représentatifs de l'ensemble du stock.

Si les prélèvements sont effectués totalement au hasard dans un carton, la variance de la moyenne $y_{i\cdot}$ du carton en question est égale à $v_A + v_B/(n_B/n_A)$. Le dernier terme de cette somme $v_B/(n_B/n_A)$ est la part de variance due aux fluctuations d'échantillonnage à l'intérieur du carton. Cette part est totalement modifiée si l'échantillonnage est systématique. Il convient alors de déterminer par une étude spécifique sa dépendance vis à vis du type d'échantillonnage systématique utilisé. Il est vraisemblable ici qu'un échantillonnage systématique avec des carottes régulièrement espacées explorant l'ensemble du carton est préférable à un échantillonnage au hasard de ces mêmes carottes. Notons enfin que l'hypothèse de proportionnalité entre le coût et le nombre de prélèvement - le coût du prélèvement intra-carton est $(n_B/n_A)c_B$ - peut être grossièrement erronée et qu'il convient donc d'examiner dans chaque cas comment varie le coût en fonction du mode d'échantillonnage adopté.

1.2 Randomisation et modèle associé

L'opération décrite pour le contrôle de réception s'appelle un sondage à deux degrés (cf J.M. Grosbras, 1987, Méthodes statistiques des sondages,

Economica). Le premier degré est le tirage du carton, le second le tirage du prélèvement dans le carton. Usuellement le nombre d'unités tirées au second degré (unités secondaires) dépend de l'unité tirée au premier degré (unité primaire). Ici, le nombre d'unités secondaires par unité primaire, c'est à dire le nombre de prélèvements par carton, est constant. En suivant un développement analogue à celui de R.Bailey (1981, A unified Approach to Design of Experiments. *J.R.Statist.Soc., Ser. A, Vol 144 214-223*), nous allons montrer qu'en pareil cas les moment d'ordre 1 et 2 (espérances, variances, covariances) des variables y_{ij} ont la même forme que dans le modèle à *effets aléatoires* défini par :

$$y_{ij} = \mu + M + A_i + B_{ij} \quad (17)$$

où les M , A_i , B_{ij} sont des effets aléatoires centrés non corrélés :

$$\begin{aligned} E(M) &= 0, & E(A_i) &= 0, & E(B_{ij}) &= 0 \\ \text{cov}(M, A_i) &= 0 & \text{cov}(M, B_{ij}) &= 0 & \text{cov}(A_i, B_{ij}) &= 0 \\ \text{cov}(A_i, A_{i'}) &= 0 & \text{pour } i &\neq i' \\ \text{cov}(B_{ij}, B_{i'j'}) &= 0 & \text{si } (i, j) &\neq (i', j') \end{aligned} \quad (18)$$

et

$$\begin{aligned} \text{var}(A_i) &= v_A & \text{var}(B_{ij}) &= v_B \\ \text{var}(M) &= - \left(\frac{v_A}{N_A} + \frac{v_B}{N_B} \right) \end{aligned} \quad (19)$$

N_A = nombre total de cartons

N_B = nombre total de prélèvements (il y a donc N_B/N_A prélèvement par carton)

Le modèle défini par (17), (18), (19) n'a pas d'existence réelle. En effet, si v_A et v_B sont positifs, la variance de M est négative. Mais les règles usuelles de calcul permettent de déduire des espérances, variances et covariance correctes

pour les y_{ij} , à savoir :

$$\begin{aligned}
E(y_{ij}) &= \mu \\
\text{cov}(y_{ij}, y_{i'j'}) &= -\left(\frac{v_A}{N_A} + \frac{v_B}{N_B}\right) && \text{si } i \neq i' \\
\text{cov}(y_{ij}, y_{ij'}) &= v_A - \left(\frac{v_A}{N_A} + \frac{v_B}{N_B}\right) && \text{si } j \neq j' \\
\text{var}(y_{ij}) &= v_A + v_B - \left(\frac{v_A}{N_A} + \frac{v_B}{N_B}\right)
\end{aligned} \tag{20}$$

On peut donc considérer (17), (18), (19) comme une simple forme mnémotechnique de (20) qui est particulièrement agréable pour les habitués de l'analyse de variance. L'intérêt de cette forme est aussi de simplifier les calculs de variance. Ainsi on déduit immédiatement de (17), (18), (19) que :

$$\text{var}(y_{..}) = \text{var}(M + A_{.} + B_{..}) = -\left(\frac{v_A}{N_A} + \frac{v_B}{N_B}\right) + \frac{v_A}{N_A} + \frac{v_B}{N_B}$$

La même formule peut être obtenue aussi à partir de (20) mais de façon nettement plus laborieuse.

La variance de l'estimateur $\hat{\mu}$ donné en (6) est donc :

$$\text{var}(\hat{\mu}) = \text{var}(y_{..} + E_{.}) = -\left(\frac{v_A}{N_A} + \frac{v_B}{N_B}\right) + \frac{v_A}{n_A} + \frac{v_B}{n_B} + \frac{v_E}{n_E}$$

Si l'on ignore le terme $v_A/N_A + v_B/N_B$ qui ne dépend pas de n_A, n_B, n_E , cette variance a la même forme que celle donnée en (7). Les effectifs n_A, n_B et n_E pour lesquels la variance est minimum ont donc la même forme que ceux obtenus précédemment avec la formule simplifiée (2). Mais la variance associée est plus petite et par conséquent le coût minimum pour obtenir une précision donnée plus faible.

Démonstration de (20)

Pour effectuer le tirage au hasard, on commence par numéroté de 1 à N_A les caisses et de 1 à N_B/N_A les emplacements possibles de prélèvement. On note $\mu + u_{\alpha\beta}$ le pourcentage de graisse à l'emplacement β de la caisse α . Le paramètre μ est par définition la moyenne sur l'ensemble des caisses et des prélèvements de ces valeurs. Donc :

$$\mu = \frac{1}{N_B} \sum_{\alpha=1}^{N_A} \sum_{\beta=1}^{N_B/N_A} (\mu + u_{\alpha\beta}) = \mu + u_{..}$$

et donc :

$$u_{..} = 0 \quad (21)$$

Les indices i et j , qui varient de 1 à n_A et de 1 à n_B/n_A respectivement font référence au numéro d'ordre du tirage effectué. Ainsi y_{ij} est le pourcentage de graisse pour le $j^{\text{ème}}$ prélèvement effectué dans le $i^{\text{ème}}$ carton tiré au hasard.

On montre sans difficulté que chacun des cartons 1 à N_A a la même probabilité $1/N_A$ d'être tiré au $i^{\text{ème}}$ tirage de carton³ et de même que chaque prélèvement 1 à N_B/N_A a la même probabilité $1/(N_B/N_A)$ d'être sélectionné au $j^{\text{ème}}$ prélèvement. La variable y_{ij} peut donc prendre chacune des N_B valeurs $\mu + u_{\alpha\beta}$ avec le même probabilité $1/N_B$. Son espérance et sa variance sont donc indépendantes du couple (i, j) considéré et de la forme :

$$E(y_{ij}) = \sum_{\alpha, \beta} \frac{1}{N_B} (\mu + u_{\alpha\beta}) = \mu \quad \text{var}(y_{ij}) = S_B = \frac{1}{N_B} \sum_{\alpha\beta} u_{\alpha\beta}^2 \quad (22)$$

On montre de même que le couple $(y_{ij}, y_{ij'})$, où $j \neq j'$, peut prendre chacun des $N_A(N_B/N_A)[N_B/N_A - 1]$ couples de valeurs $(\mu + u_{\alpha\beta}, \mu + u_{\alpha\beta'})$, où $\beta \neq \beta'$, avec la même probabilité $1/[N_B(N_B/N_A - 1)]$. La covariance $\text{cov}(y_{ij}, y_{ij'})$ est donc indépendante du triplet (i, j, j') , $j \neq j'$, considéré et de la forme :

$$\text{cov}(y_{ij}, y_{ij'}) = S_A = \frac{1}{N_B(N_B/N_A - 1)} \sum_{\substack{\alpha, \beta, \beta' \\ \beta \neq \beta'}} u_{\alpha\beta} u_{\alpha\beta'} \quad (23)$$

De même, chaque couple $(y_{ij}, y_{i'j'})$, où $i \neq i'$, peut prendre chacun des $N_A(N_A - 1)[N_B/N_A]^2$ couples de valeurs $(\mu + u_{\alpha\beta}, \mu + u_{\alpha'\beta'})$, où $\alpha \neq \alpha'$, avec une même probabilité. La covariance $\text{cov}(y_{ij}, y_{i'j'})$ est donc indépendante du quadruplet (i, i', j, j') , $i \neq i'$, considéré et de la forme :

$$\text{cov}(y_{ij}, y_{i'j'}) = S_M = \frac{1}{N_B^2(1 - 1/N_A)} \sum_{\substack{\alpha, \alpha', \beta, \beta' \\ \alpha \neq \alpha'}} u_{\alpha\beta} u_{\alpha'\beta'} \quad (24)$$

3. cette probabilité est le rapport entre le nombre $(N_A - 1) \cdots (N_A - n_A + 1)$ d'arrangements de n_A cartons parmi les N_A qui contiennent le dit carton en position i , et le nombre total $N_A(N_A - 1) \cdots (N_A - n_A + 1)$ d'arrangements.

Notons que

$$N_B S_B + N_B \left(\frac{N_B}{N_A} - 1 \right) S_A + N_B^2 \left(1 - \frac{1}{N_A} \right) S_M = \sum_{\alpha, \alpha', \beta, \beta'} u_{\alpha\beta} u_{\alpha'\beta'} = \left(\sum_{\alpha, \beta} u_{\alpha\beta} \right)^2.$$

Le dernier terme est nul et on a donc

$$N_B S_B + N_B \left(\frac{N_B}{N_A} - 1 \right) S_A + N_B^2 \left(1 - \frac{1}{N_A} \right) S_M = 0 \quad (25)$$

ou encore

$$\frac{S_B - S_A}{N_B} + \frac{S_A - S_M}{N_A} + S_M = 0 \quad (26)$$

Posant alors :

$$v_B = S_B - S_A, \quad v_A = S_A - S_M \quad (27)$$

on déduit facilement (20) des égalités (22), (23), (24), (26).

Bardin et Azaïs (1990, une hypothèse minimale pour une théorie des plans d'expériences randomisés, RSA, Vol 38, 2, 21-41) utilisent des résultats élémentaires de la théorie des groupes de permutations pour démontrer des résultats analogues dans un cadre beaucoup plus général. Leur article s'applique au cas où $n_A = N_A$, $n_B = N_B$, c'est à dire au cas d'un tirage exhaustif. Mais on peut sans difficulté l'étendre au cas d'un tirage non exhaustif en imaginant que l'on poursuit le tirage au hasard au-delà du n_A ème carton et du (n_B/n_A) ème prélèvement dans chaque carton, pour aboutir in fine à un tirage exhaustif. La distribution des y_{ij} correspondant aux n_B premiers prélèvements ne dépend pas des résultats qui seraient obtenus dans les prélèvements ultérieurs et cette distribution est donc identique à ce qu'elle serait si le tirage était exhaustif.

Notons que :

$$N_B S_B + N_B \left(\frac{N_B}{N_A} - 1 \right) S_A = \sum_{\alpha} \left(\sum_{\beta} u_{\alpha\beta} \right)^2 = \left(\frac{N_B}{N_A} \right)^2 \sum_{\alpha} u_{\alpha}^2. \quad (28)$$

De (22), (25) et (28) on déduit

$$v_B = S_B - S_A = \frac{\sum_{\alpha\beta} u_{\alpha\beta}^2 - \sum_{\alpha\beta} u_{\alpha}^2}{N_A(N_B/N_A - 1)} \quad (29)$$

$$\frac{N_B}{N_A}v_A + v_B = \frac{\sum_{\alpha\beta} u_{\alpha}^2}{(N_A - 1)} \quad (30)$$

Les quantités apparaissant à droite de (29) et (30) sont respectivement les carrés moyens intra et inter-carton calculés sur l'ensemble des $u_{\alpha\beta}$.

Ces résultats peuvent être étendus à certaines des situations plus générales considérées par Bardin, Azaïs (1990).

1.3 Estimation des paramètres v_A, v_B, v_E

La recherche de n_A, n_B et n_E menée au paragraphe 1.1 suppose connues les variances v_A, v_B et v_E . En pratique, ces variances doivent avoir été estimées lors d'une expérience antérieure au cours de laquelle on a effectué les mesures sur chaque prélèvement. Le modèle associé à cette expérience préliminaire est :

$$Y_{ijk} = \mu_0 + A_i + B_{ij} + E_{ijk} \quad (31)$$

où

$$\begin{aligned} \text{var}(A_i) &= v_A \\ \text{var}(B_{ij}) &= v_B \\ \text{var}(E_{ijk}) &= v_E \end{aligned} \quad (32)$$

et les aléatoires A_i, B_{ij}, E_{ijk} sont supposées centrées et de covariance nulle. On espère que, même si la moyenne μ pour le sondage en cours est différente de la moyenne μ_0 de l'expérience préliminaire, les ordres de grandeur des variances ne sont pas trop différents et que l'on peut s'appuyer sur les estimations $\hat{v}_A, \hat{v}_B, \hat{v}_E$ obtenues lors de l'expérience préliminaire pour planifier le sondage.

Soient I le nombre de cartons, J le nombre de prélèvements par carton et K le nombre de mesures par prélèvement dans l'expérience préliminaire. Une analyse de variance permet d'obtenir les carrés moyens CM_A, CM_B et

CM_E associés aux trois effets aléatoires A_i , B_{ij} et E_{ijk} :

$$\begin{aligned} CM_A &= \frac{\sum_{ijk}(Y_{i..} - Y_{...})^2}{I - 1} = \frac{JK}{I - 1} \sum_i (Y_{i..} - Y_{...})^2 \\ CM_B &= \frac{\sum_{ijk}(Y_{ij.} - Y_{...})^2}{I(J - 1)} = \frac{K}{I(J - 1)} \sum_{ij} (Y_{ij.} - Y_{...})^2 \\ CM_E &= \frac{\sum_{ijk}(Y_{ijk} - Y_{ij.})^2}{IJ(K - 1)} \end{aligned} \quad (33)$$

Les espérances de ces carrés moyens sont données par les formules classiques :

$$\begin{aligned} E(CM_E) &= v_E \\ E(CM_B) &= r_B v_B + v_E \\ E(CM_A) &= r_A v_A + r_B v_B + v_E \end{aligned} \quad (34)$$

où r_A , r_B sont respectivement les nombres de mesures par caisse et par prélèvement, soit :

$$r_A = JK \quad r_B = K \quad (35)$$

On déduit de (34) et (35) les formules suivantes :

$$\begin{aligned} v_E &= E(CM_E) \\ v_B &= (E(CM_B) - E(CM_E))/K \\ v_A &= (E(CM_A) - E(CM_B))/JK \end{aligned}$$

qui conduisent à prendre comme estimations de v_E , v_B , v_A

$$\begin{aligned} \hat{v}_E &= CM_E \\ \hat{v}_B &= (CM_B - CM_E)/K \\ \hat{v}_A &= (CM_A - CM_B)/JK \end{aligned} \quad (36)$$

Remarques

- On peut se demander si la variance v_E des mesures effectuées à partir d'un seul prélèvement est identique à la variance des mesures effectuées, comme c'est le cas dans le sondage, à partir du mélange convenablement homogénéisé de n_B prélèvements. Si on dispose, en plus des résultats de

l'expérience préliminaire, d'une série de mesures effectuées à partir d'un tel mélange, on répond à cette question en comparant le \hat{v}_E obtenu par (36) et la variance de cette série de mesures (par un test de Snedecor pour la comparaison de deux variances qui est décrit par exemple dans l'ouvrage collectif du CEA, 1978, Statistique appliquée à l'exploitation des mesures, tome II). En cas de différence significative, c'est cette dernière variance que l'on retiendra comme variance de la mesure dans les sondages.

- Si l'expérience préliminaire est convenablement randomisée, on a pour cette expérience un modèle analogue à (20), où les variances sont données par des formules telles que (29), (30). Les quantités N_A , N_B , $u_{\alpha\beta}$ sont celles qui correspondent à l'ensemble des prélèvements possibles pour cette expérience préliminaire et sont distinctes des quantités analogues utilisées pour les sondages ultérieurs.

Si les variabilités inter et intra-carton sont du même ordre de grandeur dans l'expérience préliminaire et lors du sondage, les termes v_B et v_A définis à partir des $u_{\alpha\beta}$ de l'expérience préliminaire sont voisins des quantités homologues pour le stock à sonder. Pour justifier statistiquement cette affirmation, on est obligé de considérer les $u_{\alpha\beta}$ comme des réalisations de quantités aléatoires $U_{\alpha\beta}$ ayant une même distribution dans l'expérience préliminaire et dans le stock à sonder :

$$U_{\alpha\beta} = \mathcal{A}_\alpha + \mathcal{B}_{\alpha\beta}$$

où les \mathcal{A}_α , $\mathcal{B}_{\alpha\beta}$ sont des aléatoires centrés non corrélés de variances :

$$\text{var}(\mathcal{A}_\alpha) = v_A \quad \text{var}(\mathcal{B}_{\alpha\beta}) = v_B$$

Les quantités v_A et v_B déduites de $u_{\alpha\beta}$ de l'expérience préliminaire et les quantités homologues déduites des $u_{\alpha\beta}$ du stock à sonder sont alors les estimateurs usuels des paramètres communs v_A et v_B .

2 Etude d'un facteur à 2 niveaux

2.1 Exemple et modèle

Supposons pour fixer les idées qu'on ait à comparer deux régimes alimentaires pour porcelets, du point de vue de l'accroissement de poids sur une période déterminée. Le *facteur* dont on étudie l'influence est ici le "Régime

Alimentaire". Il a deux *niveaux*, ou *modalités*, que l'on numérote de façon habituelle 1 et -1 . On note n_1 et n_{-1} le nombre d'animaux qui ont reçu les régimes 1 et -1 respectivement. L'accroissement de poids, qui est la *variable expliquée* ou *réponse*, est noté y .

La réponse du $j^{\text{ème}}$ animal qui a reçu le régime i est noté y_{ij} . On suppose que y_{ij} suit le modèle :

$$\begin{aligned} E(y_{ij}) &= \mu_i \\ \text{var}(y_{ij}) &= \sigma_i^2 \\ \text{cov}(y_{ij}, y_{i'j'}) &= 0 \quad \text{si } i \neq i' \text{ ou } j \neq j' \end{aligned}$$

que l'on peut aussi écrire, en posant $\varepsilon_{ij} = y_{ij} - \mu_i$, sous la forme

$$\left. \begin{aligned} y_{ij} &= \mu_i + \varepsilon_{ij} & E(\varepsilon_{ij}) &= 0 \\ \text{var}(\varepsilon_{ij}) &= \sigma_i^2 \\ \text{cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) &= 0 & \text{si } i \neq i' \text{ ou } j \neq j' \end{aligned} \right\} \quad (37)$$

L'espérance μ_i de la réponse est appelée *réponse théorique* pour le régime i , ou encore *moyenne théorique*, ou s'il n'y a pas d'ambiguïté, *moyenne* tout court. Le terme ε_{ij} est appelé *erreur*, bien qu'il s'agisse plutôt ici d'un effet propre au $j^{\text{ème}}$ animal qui a reçu le traitement i . On fait souvent l'hypothèse, dite *d'homoscédasticité*, que les variances σ_i^2 de ces erreurs ne dépendent pas du régime i . Si σ^2 est leur valeur commune, on a alors:

$$\text{var}(\varepsilon_{ij}) = \sigma^2 \quad (38)$$

Cette hypothèse est généralement réaliste, mais il arrive qu'elle soit grossièrement erronée. Ainsi si l'un des régimes contient une substance susceptible de produire, chez certains animaux seulement, une perte d'appétit, ce régime engendrera une diminution de l'accroissement de poids moyen mais en même temps une forte augmentation de la variabilité.

L'hypothèse de nullité des covariances est souvent considérée comme une conséquence d'une hypothèse plus forte d'indépendance entre les accroissements de poids y_{ij} des différents animaux.

2.2 Rappel sur l'estimation des paramètres

Les estimateurs usuels des paramètres μ_i , σ_i^2 sont

$$\hat{\mu}_i = y_{i\cdot} \quad \hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_{j,j'} (y_{ij} - y_{i\cdot})^2$$

où le point indique qu'on a effectué la moyenne sur l'indice correspondant :

$$y_{i\bullet} = \frac{y_{11} + \dots + y_{1n_i}}{n_i}$$

Les estimateurs $\hat{\mu}_1$ et $\hat{\mu}_{-1}$ sont non corrélés et ont pour variance :

$$\text{var}(\hat{\mu}_i) = \frac{\sigma_i^2}{n_i}$$

La différence $\mu_1 - \mu_{-1}$ est estimée par :

$$\hat{\mu}_1 - \hat{\mu}_{-1} = y_{1\bullet} - y_{-1\bullet}$$

La variance correspondante

$$\text{var}(\hat{\mu}_1 - \hat{\mu}_{-1}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_{-1}^2}{n_{-1}}$$

est estimée par

$$\widehat{\text{var}}(\hat{\mu}_1 - \hat{\mu}_{-1}) = \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_{-1}^2}{n_{-1}}$$

Les estimations des paramètres μ_i et les variances de ces estimations peuvent être retrouvées à partir des formules matricielles usuelles. Nous considérons à titre d'exemple le cas où $n_1 = 2$, $n_{-1} = 1$. Le modèle (37) s'écrit

$$\begin{array}{llllll} y_{11} & = & \mu_1 & + & \varepsilon_{11} & \text{var}(\varepsilon_{11}) & = & \sigma_1^2 \\ y_{12} & = & \mu_1 & + & \varepsilon_{12} & \text{var}(\varepsilon_{11}) & = & \sigma_1^2 \\ y_{-11} & = & \mu_{-1} & + & \varepsilon_{-11} & \text{var}(\varepsilon_{-11}) & = & \sigma_{-1}^2 \end{array} \quad \text{cov}(\varepsilon_{11}, \varepsilon_{-11}) = 0$$

où encore sous forme matricielle

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{-11} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} \mu_1 \\ \mu_{-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{-11} \end{bmatrix} \quad \text{cov} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{-11} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_{-1}^2 \end{bmatrix}$$

soit

$$\mathbf{y} = X\boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad \text{cov}(\boldsymbol{\varepsilon}) = \Sigma$$

avec

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{-11} \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_{-1} \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{-11} \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_{-1}^2 \end{bmatrix}$$

Les formules matricielles donnant $\hat{\boldsymbol{\mu}}$ et sa matrice de covariance sont:

$$\hat{\boldsymbol{\mu}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\mathbf{y} \quad \text{cov}(\hat{\boldsymbol{\mu}}) = (X'\Sigma^{-1}X)^{-1}$$

Le calcul donne :

$$X'\Sigma^{-1}X = \begin{bmatrix} \frac{n_1}{\sigma_1^2} & 0 \\ 0 & \frac{n_{-1}}{\sigma_{-1}^2} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_{-1}^2} \end{bmatrix}$$

d'où

$$\text{cov}(\hat{\boldsymbol{\mu}}) = (X'\Sigma^{-1}X)^{-1} = \begin{bmatrix} \frac{\sigma_1^2}{n_1} & 0 \\ 0 & \frac{\sigma_{-1}^2}{n_{-1}} \end{bmatrix} = \begin{bmatrix} \frac{\sigma_1^2}{2} & 0 \\ 0 & \frac{\sigma_{-1}^2}{1} \end{bmatrix}$$

$$X'\Sigma^{-1}\mathbf{y} = \begin{bmatrix} \frac{n_1 y_{1\cdot}}{\sigma_1^2} \\ \frac{n_{-1} y_{-1\cdot}}{\sigma_{-1}^2} \end{bmatrix} \quad \hat{\boldsymbol{\mu}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\mathbf{y} = \begin{bmatrix} y_{1\cdot} \\ y_{-1\cdot} \end{bmatrix}$$

Rappelons que les estimateurs ainsi trouvés sont de variance minimum parmi les estimateurs linéaires sans biais de \mathbf{y} .

2.3 Randomisation⁴

Le modèle (37) peut-être justifié en prenant en compte le tirage au hasard, ou *randomisation*, qui est normalement effectué pour choisir les animaux qui reçoivent le régime 1 et ceux qui reçoivent le régime -1 . L'animal sélectionné au $j^{\text{ème}}$ tirage pour le régime i peut être n'importe lequel des animaux susceptibles de participer à l'expérience. On repère ces derniers, au nombre de N , par un indice α variant de 1 à N et on note $u_{i\alpha}$ l'accroissement de poids qu'aurait l'animal α si il était soumis au régime i . Alors y_{ij} peut prendre n'importe laquelle des valeurs $u_{i\alpha}$ avec une même probabilité $1/N$.

4. Ce paragraphe plus technique peut être laissé de coté dans une première lecture

Ceci montre que les y_{ij} , pour j variant de 1 à n_i , ont même distribution, donc même moyenne μ_i et même variance θ_i^2 , à savoir :

$$\mu_i = \frac{1}{N} \sum_{\alpha=1}^N u_{i\alpha} \quad \theta_i^2 = \frac{1}{N} \sum_{\alpha=1}^N (u_{i\alpha} - \mu_i)^2 \quad (39)$$

A deux tirages distincts correspondent des animaux distincts. Le couple $(y_{ij}, y_{i'j'})$ correspondant à ces deux tirages peut donc prendre n'importe lequel des $N(N-1)$ couples de valeurs $(u_{i\alpha}, u_{i'\beta})$ où $\alpha \neq \beta$. Sa distribution ne dépend donc pas de j et j' , et il en est de même de la covariance $c_{ii'}$ entre y_{ij} et $y_{i'j'}$, qui est :

$$\begin{aligned} c_{ii'} &= \text{cov}(y_{ij}, y_{i'j'}) = \frac{1}{N(N-1)} \sum_{\substack{\alpha, \beta \\ \alpha \neq \beta}} (u_{i\alpha} - \mu_i)(u_{i'\beta} - \mu_{i'}) \\ &= \frac{1}{N(N-1)} \left[\left(\sum_{\alpha} (u_{i\alpha} - \mu_i) \right) \left(\sum_{\beta} (u_{i'\beta} - \mu_{i'}) \right) - \sum_{\alpha} (u_{i\alpha} - \mu_i)(u_{i'\alpha} - \mu_{i'}) \right] \\ &= -\frac{1}{N(N-1)} \sum_{\alpha=1}^N (u_{i\alpha} - \mu_i)(u_{i'\alpha} - \mu_{i'}) \end{aligned}$$

On a donc :

$$\text{cov}(y_{ij}, y_{ij'}) = c_{ii} = -\frac{\theta_i^2}{N-1} \quad \text{pour } j \neq j' \quad (40)$$

$$\text{cov}(y_{1j}, y_{-1j'}) = c_{1,-1} = -\frac{C}{N-1} \quad (41)$$

où

$$C = \frac{1}{N} \sum_{\alpha=1}^N (u_{1\alpha} - \mu_1)(u_{-1\alpha} - \mu_{-1}) \quad (42)$$

est la covariance entre les réponses aux régimes 1 et -1 des animaux 1 à N .

Les espérances, variances et covariances sont identiques à celles qu'on obtient à partir du modèle :

$$y_{ij} = \mu_i + M + M_i + \varepsilon_{ij} \quad (43)$$

où les aléatoires centrées M, M_i, ε_{ij} sont non corrélées et ont des variances définies par

$$\begin{aligned} \text{var}(M) &= -\frac{C}{N-1} \\ \text{var}(M) + \text{var}(M_i) &= -\frac{\theta_i^2}{N-1} \\ \text{var}(M) + \text{var}(M_i) + \text{var}(\varepsilon_{ij}) &= \theta_i^2 \end{aligned} \quad (44)$$

soit telles que

$$\begin{aligned} \text{var}(M) &= -\frac{C}{N-1} \\ \text{var}(M_i) &= -\frac{\theta_i^2 - C}{N-1} \\ \text{var}(\varepsilon_{ij}) &= \frac{N}{N-1} \theta_i^2. \end{aligned} \quad (45)$$

Ce modèle n'a pas d'existence réelle, puisque certains des termes qui y figurent peuvent avoir une variance négative, mais il est commode pour obtenir les variances et covariances de combinaisons linéaires des y_{ij} . On en déduit par exemple que les trois ensembles de combinaisons linéaires suivants :

$$\begin{aligned} 1. \quad y_{1j} - y_{1\cdot} &= \varepsilon_{1j} - \varepsilon_{1\cdot} \\ 2. \quad y_{-1,j} - y_{-1\cdot} &= \varepsilon_{-1j} - \varepsilon_{-1\cdot} \\ 3. \quad \begin{cases} y_{1\cdot} &= \mu_1 + M + M_1 + \varepsilon_{1\cdot} \\ y_{-1\cdot} &= \mu_{-1} + M + M_{-1} + \varepsilon_{-1\cdot} \end{cases} \end{aligned}$$

sont non corrélés. A partir des deux premiers, on peut estimer sans biais θ_1^2 et θ_{-1}^2 :

$$\frac{N}{N-1} \hat{\theta}_i^2 = \widehat{\text{var}(\varepsilon_{ij})} = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - y_{i\cdot})^2. \quad (46)$$

Du troisième on déduit des estimations sans biais de μ_1 et μ_{-1} et les variances de ces estimations :

$$\begin{aligned} \hat{\mu}_i &= y_{i\cdot} \\ \text{var}(\hat{\mu}_i) &= \frac{N}{N-1} \theta_i^2 \left(\frac{1}{n_i} - \frac{1}{N} \right) \end{aligned} \quad (47)$$

L'estimation de cette dernière variance s'obtient en remplaçant θ_i^2 par son estimation $\hat{\theta}_i^2$:

$$\widehat{\text{var}}(\hat{\mu}_i) = \frac{N}{N-1} \hat{\theta}_i^2 \left(\frac{1}{n_i} - \frac{1}{N} \right) \quad (48)$$

La différence $\mu_1 - \mu_{-1}$ est estimée sans biais par :

$$\hat{\mu}_1 - \hat{\mu}_{-1} = y_{1\bullet} - y_{-1\bullet} = \mu_1 - \mu_{-1} + M_1 - M_{-1} + \varepsilon_{1\bullet} - \varepsilon_{-1\bullet} \quad (49)$$

La variance de cette différence est:

$$\text{var}(\hat{\mu}_1 - \hat{\mu}_{-1}) = -\frac{\theta_1^2 + \theta_{-1}^2 - 2C}{N-1} + \frac{N}{N-1} \left(\frac{\theta_1^2}{n_1} + \frac{\theta_{-1}^2}{n_{-1}} \right) \quad (50)$$

soit

$$\text{var}(\hat{\mu}_1 - \hat{\mu}_{-1}) = \frac{N}{N-1} \theta_1^2 \left(\frac{1}{n_1} - \frac{1}{N} \right) + \frac{N}{N-1} \theta_{-1}^2 \left(\frac{1}{n_{-1}} - \frac{1}{N} \right) + \frac{2C}{N-1} \quad (51)$$

Le troisième terme $2C/(N-1)$ provient de la covariance créée entre $y_{1\bullet}$ et $y_{-1\bullet}$ par le terme M commun. On n'a aucun moyen d'estimer cette covariance. Mais on peut remarquer les faits suivants:

- Si n_1 et n_{-1} sont petits devant $N/2$, le terme $\frac{\theta_1^2 + \theta_{-1}^2 - 2C}{N-1}$ dans l'expression (50) de la variance est négligeable. En appliquant l'inégalité de Schwartz à l'égalité (42) de définition de C , on obtient en effet

$$2|C| \leq 2\theta_1\theta_{-1} \leq \theta_1^2 + \theta_{-1}^2$$

d'où

$$\frac{\theta_1^2 + \theta_{-1}^2 - 2C}{N-1} \leq \frac{N}{N-1} \frac{2(\theta_1^2 + \theta_{-1}^2)}{N} \ll \frac{N}{N-1} \left(\frac{\theta_1^2}{n_1} + \frac{\theta_{-1}^2}{n_{-1}} \right)$$

On a alors

$$\text{var}(\hat{\mu}_1 - \hat{\mu}_{-1}) \simeq \frac{N}{N-1} \left(\frac{\theta_1^2}{n_1} + \frac{\theta_{-1}^2}{n_{-1}} \right) \quad (52)$$

Du point de vue estimation de la différence $\mu_1 - \mu_{-1}$, et estimation de la variance de cette différence, tout se passe donc comme si on avait le modèle suivant d'analyse de la variance, où les erreurs ε_{ij} , non corrélées deux à deux, ont une variance $\sigma_i^2 = \frac{N}{N-1} \theta_i^2$ qui peut dépendre de i :

$$y_{ij} = \mu_i + \varepsilon_{ij} \quad \text{var}(\varepsilon_{ij}) = \sigma_i^2 \quad (53)$$

- Supposons que la différence $u_{1\alpha} - u_{-1\alpha}$ soit la même pour tous les animaux. Les réponses $u_{i\alpha}$ peuvent se mettre sous la forme additive

$$u_{i\alpha} = \mu_i + \beta_\alpha \quad (54)$$

avec $\sum_\alpha \beta_\alpha = 0$. On a alors $\theta_1^2 = \theta_{-1}^2 = C = (\sum \beta_\alpha^2) / N$, d'où $\theta_1^2 + \theta_{-1}^2 - 2C = 0$ et la variance de $\hat{\mu}_1 - \hat{\mu}_{-1}$ est donnée par l'expression (52) dans laquelle on peut remplacer $\theta_1 = \theta_{-1}$ par leur valeur commune θ .

Tout se passe donc comme si le modèle (53) dans lequel on remplace σ_i^2 par $\sigma^2 = \frac{N}{N-1}\theta^2$ était valide.

- Dans tous les cas de figure, la variance de $\hat{\mu}_1 - \hat{\mu}_{-1}$ est majorée par l'expression $\frac{N}{N-1}(\frac{\theta_1^2}{n_1} + \frac{\theta_{-1}^2}{n_{-1}})$ donnée en (52).

On peut donc justifier l'emploi d'un modèle d'analyse de variance tel que (53) par la prise en compte de la randomisation. Si en outre les réponses potentielles $u_{i\alpha}$ suivent un modèle tel que (54), les variances σ_i^2 figurant dans (53) sont égales et on peut écrire $\text{var}(\varepsilon_{ij}) = \sigma^2$ où σ^2 est la valeur commune de σ_1^2 et σ_{-1}^2 .

En pratique, la population pour laquelle on veut estimer la différence moyenne $\mu_1 - \mu_{-1}$ est beaucoup plus vaste que celle au sein de laquelle s'effectue effectivement le choix des animaux de l'expérience. On n'a donc pas la possibilité de randomiser librement pour obtenir un échantillon représentatif. Par ailleurs, l'expérimentation se passe dans des conditions souvent fort différentes des conditions d'application des résultats car on cherche à réduire au maximum la variabilité pour détecter une différence aussi petite que possible. Il s'en suit qu'on empêche la fluctuation de certaines sources de variation qui interagissent avec le facteur étudié et pourraient modifier sensiblement les résultats. Pour remédier à cela, le mieux est généralement de faire varier de façon contrôlée dans un plan d'expérience multifactoriel celles de ces autres sources de variation qui sont clairement identifiées. On obtient ainsi des résultats beaucoup plus représentatifs, sans perdre de précision sur l'estimation de l'effet de facteurs concernés. Mais il faut garder à l'esprit qu'on ne peut pas tout contrôler et que l'on risque toujours en se plaçant dans d'aussi bonnes conditions d'expérimentation que possible de limiter la portée des résultats obtenus.

2.4 Plan optimal

L'estimation des moindres carrés de μ_i dans le modèle (37) est $y_{i\cdot}$. L'estimation de $\mu_1 - \mu_{-1}$ est donc :

$$\hat{\mu}_1 - \hat{\mu}_{-1} = y_{1\cdot} - y_{-1\cdot} \quad (55)$$

et sa variance est :

$$\text{var}(\hat{\mu}_1 - \hat{\mu}_{-1}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_{-1}^2}{n_{-1}} \quad (56)$$

On suppose dans ce qui suit que σ_1^2 et σ_{-1}^2 sont connus. En pratique, sauf si on a des raisons particulières de supposer que $\sigma_1 \neq \sigma_{-1}$, on fait l'hypothèse d'homoscédaticité $\sigma_1 = \sigma_{-1}$ et on utilise comme valeur commune σ de cet écart-type une valeur antérieurement estimée à partir de données similaires.

Soit $c_1 n_1 + c_{-1} n_{-1}$ le coût associé au plan défini par les effectifs n_1 et n_{-1} . La méthode des multiplicateurs de Lagrange montre que le minimum de la variance (56) sous la contrainte $c_1 n_1 + c_{-1} n_{-1} \leq c$ est obtenu lorsque la contrainte est active (c'est à dire qu'il y a égalité) et que n_1, n_{-1} sont proportionnels respectivement à $\sigma_1/\sqrt{c_1}$ et $\sigma_{-1}/\sqrt{c_{-1}}$. La variance minimum est donc obtenue lorsque:

$$n_1 = \lambda \frac{\sigma_1}{\sqrt{c_1}}, \quad n_{-1} = \lambda \frac{\sigma_{-1}}{\sqrt{c_{-1}}} \quad \text{avec } \lambda = \frac{c}{\sigma \sqrt{c_1} + \sigma_{-1} \sqrt{c_{-1}}} \quad (57)$$

La variance minimale correspondante est:

$$v_m(c) = \frac{(\sigma_1 \sqrt{c_1} + \sigma_{-1} \sqrt{c_{-1}})^2}{c} \quad (58)$$

Considérons en particulier le cas où il y a homoscédaticité et où les coûts des deux régimes sont identiques:

$$\sigma_1 = \sigma_{-1}, \quad c_1 = c_{-1} \quad (59)$$

On peut supposer que $c_1 = c_{-1} = 1$. Cela revient en effet à chiffrer le coût en nombre d'animaux utilisés. Notons n , au lieu de c , le nombre maximum d'animaux utilisables, et σ l'écart-type commun. On trouve que la variance minimum:

$$v_m(n) = \frac{4\sigma^2}{n} \quad (60)$$

est obtenue lorsque

$$n_1 = n_{-1} = \frac{n}{2} \quad (61)$$

On est souvent obligé avant l'expérience de faire, faute de mieux, l'hypothèse d'égalité des variances. Si cette hypothèse est fausse, les effectifs obtenus ne sont pas optimaux et la variance correspondante de $\hat{\mu}_1 - \hat{\mu}_{-1}$ est supérieure à celle qu'on aurait obtenue si on avait deviné correctement le quotient σ_1/σ_{-1} et choisit les effectifs en conséquence.

Etudions cette perte de précision dans le cas où $c_1 = c_{-1} = 1$. Faute de connaître le ratio σ_1/σ_{-1} on prend donc le plan défini par (61) avec des effectifs égaux à $n/2$. La variance d'estimation (56) est alors $2(\sigma_1^2 + \sigma_{-1}^2)/n$ alors que la variance minimum que l'on aurait eue si on avait deviné correctement le rapport σ_1/σ_{-1} est $(\sigma_1 + \sigma_{-1})^2/n$. Le rapport de ces variances est :

$$\frac{2(\sigma_1^2 + \sigma_{-1}^2)}{(\sigma_1 + \sigma_{-1})^2} = \frac{(\sigma_1 + \sigma_{-1})^2 + (\sigma_1 - \sigma_{-1})^2}{(\sigma_1 + \sigma_{-1})^2} = 1 + \frac{(\frac{\sigma_1}{\sigma_{-1}} - 1)^2}{(\frac{\sigma_1}{\sigma_{-1}} + 1)^2}$$

Le tableau 1 donne ce rapport pour quelque valeur du quotient σ_1/σ_{-1} . On y a reporté également le rapport, plus parlant, des écarts types correspondants :

σ_1/σ_{-1}	1	2	3	4	5
Rapport des variances	1	1.11	1.25	1.36	1.44
Rapport des écart-types	1	1.05	1.12	1.17	1.20

TAB. 1 – *Ecart à l'optimalité du plan à effectifs égaux dans le cas de variances inégales*

2.5 Efficacité et estimabilité

On suppose toujours vérifiée l'hypothèse (59) d'homoscédaticité et d'égalité des coûts. La recherche d'une précision maximum dans la comparaison entre les deux régimes conduit à soumettre le même nombre d'animaux à chacun des deux régimes. Mais on peut se demander quelle est la perte de précision lorsque pour une raison quelconque les effectifs n_1 et n_{-1} ne sont pas égaux. On introduit pour cela l'*efficacité* (généralement dite factorielle) du plan concerné.

C'est le rapport entre la variance du plan optimal pour $n = n_1 + n_{-1}$ animaux et la variance du plan considéré:

$$\text{efficacité} = \frac{\frac{4}{n}\sigma^2}{\left(\frac{1}{n_1} + \frac{1}{n-n_1}\right)\sigma^2} = 4\frac{n_1}{n} \left(1 - \frac{n_1}{n}\right)$$

On a reporté cette efficacité pour quelques valeurs du rapport n_1/n dans le tableau 2. La quantité $1/\sqrt{\text{efficacité}}$ figurant dans la 3ème ligne de ce tableau donne le coefficient par lequel il faut multiplier l'écart-type du dispositif optimal pour obtenir l'écart-type du plan considéré.

n_1/n	0	1/10	2/10	3/10	4/10	5/10
efficacité	0	0.36	0.64	0.84	.96	1
$1/\sqrt{\text{efficacité}}$	∞	1.67	1.25	1.09	1.02	1

TAB. 2 – *Efficacité dans le plan pour 1 facteur*

Il faut un déséquilibre assez fort entre les effectifs n_1 et n_{-1} pour que l'écart-type soit augmenté de façon notable. Dans le cas où $n_1 = 0$, on ne peut estimer ni μ_1 , ni a fortiori $\mu_1 - \mu_{-1}$, puisqu'il n'y a pas de données pour le régime 1. Ces deux paramètres sont dits *non estimables*. Cette non estimabilité se traduit par une variance infinie et une efficacité nulle.

3 Etude de 2 facteurs à 2 niveaux

3.1 Exemples

On considère maintenant un plan d'expérience étudiant l'influence de deux facteurs A et B à deux niveaux chacun, sur une variable y qui est appelée la *réponse*. Par exemple dans une fabrication de fromages, on compare les temps de prise obtenus avec deux présures sur deux laits d'origines distinctes. Ou encore on étudie le taux de croissance d'une bactérie au pH 7.5 et 8.5 et aux températures 20°C et 30°C.

- Dans le premier cas la réponse y est le temps de prise et les facteurs A et B sont le type de présure et l'origine du lait. Les *traitements*

sont les combinaisons des niveaux de A et B . Ils sont donnés à gauche du tableau 3 avec la numérotation usuelle 1, -1 pour les facteurs à 2 niveaux.

- Dans le second cas, la réponse y est le taux de croissance et A et B sont respectivement le pH et la température. les 4 traitements sont également indiqués à droite du tableau 3 avec mention entre parenthèse du niveau explicite associé au niveau codé 1 ou -1 . Les facteurs sont ici quantitatifs, mais on n'en tiendra pas compte dans ce paragraphe.

type de présure	origine du lait		pH	température
-1	-1	traitement 1	-1 (7.5)	-1 ($20^{\circ}C$)
-1	1	traitement 2	-1 (7.5)	1 ($30^{\circ}C$)
1	-1	traitement 3	1 (8.5)	-1 ($20^{\circ}C$)
1	1	traitement 4	1 (8.5)	1 ($30^{\circ}C$)
Exemple des présures			Exemple des bactéries	

TAB. 3 – *Traitements dans des plans pour 2 facteurs à 2 niveaux*

On note n_{ij} le nombre d'unités expérimentales qui reçoivent le traitement ij . Par unité expérimentale, il faut entendre un fromage fabriqué à partir d'une présure et d'un lait donné dans le premier exemple, où une fermentation réalisée à un pH initial et une température déterminée dans le second exemple. Ce nombre n_{ij} est aussi appelé nombre de répétitions du traitement ij . Le plan d'expérience est déterminé par la donnée des nombres de répétitions associés aux 4 traitements. On le présente dans la suite sous la forme du tableau 4 .

		$B(j)$	
		-1	1
$A(i)$	-1	$n_{-1,-1}$	$n_{-1,1}$
	1	$n_{1,-1}$	$n_{1,1}$

TAB. 4 – *Le plan d'expérience*

3.2 Modèle et effets factoriels

On note y_{ijk} la réponse pour la $k^{\text{ème}}$ répétition du traitement ij . On suppose que les réponses y_{ijk} suivent le modèle:

$$\begin{aligned} y_{ijk} &= \mu_{ij} + \varepsilon_{ijk} & E(\varepsilon_{ijk}) &= 0 \\ \text{var}(\varepsilon_{ijk}) &= \sigma_{ij}^2 \\ \text{cov}(\varepsilon_{ijk}, \varepsilon_{i'j'k'}) &= 0 & \text{si } i \neq i' \text{ ou } j \neq j' \text{ ou } k \neq k' \end{aligned} \quad (62)$$

On ajoute parfois l'hypothèse d'homoscédaticité qui permet d'écrire:

$$\text{var}(\varepsilon_{ijk}) = \sigma^2 \quad (63)$$

Dans une telle étude, on s'intéresse bien sûr aux moyennes théoriques μ_{ij} , mais aussi et surtout aux comparaisons que l'on peut faire entre ces moyennes, en particulier à ce que l'on appelle les *effets factoriels*. Ceux-ci sont les effets principaux des facteurs A et B , notés $e(A)$, $e(B)$ et l'interaction $e(AB)$. Ils sont définis par :

$$\begin{aligned} e(A) &= \frac{1}{2}(\mu_{1\cdot} - \mu_{-1\cdot}) \\ e(B) &= \frac{1}{2}(\mu_{\cdot 1} - \mu_{\cdot -1}) \\ e(AB) &= \frac{1}{4}[(\mu_{11} - \mu_{1-1}) - (\mu_{-11} - \mu_{-1-1})] \end{aligned} \quad (64)$$

La donnée de ces trois effets factoriels et de la moyenne générale, notée $e(\mathcal{K})$, définie par

$$e(\mathcal{K}) = \mu_{\cdot\cdot} \quad (65)$$

est équivalente à la donnée des 4 réponses μ_{-1-1} , μ_{1-1} , μ_{-11} , μ_{11} . Les définitions (64) et (65) peuvent en effet se réécrire sous la forme:

$$\begin{aligned} 4e(\mathcal{K}) &= \mu_{-1-1} + \mu_{-11} + \mu_{1-1} + \mu_{11} \\ 4e(A) &= -\mu_{-1-1} - \mu_{-11} + \mu_{1-1} + \mu_{11} \\ 4e(B) &= -\mu_{-1-1} + \mu_{-11} - \mu_{1-1} + \mu_{11} \\ 4e(AB) &= \mu_{-1-1} - \mu_{-11} - \mu_{1-1} + \mu_{11} \end{aligned} \quad (66)$$

d'où l'on déduit immédiatement :

$$\begin{aligned} \mu_{-1-1} &= e(\mathcal{K}) - e(A) - e(B) + e(AB) \\ \mu_{-1\ 1} &= e(\mathcal{K}) - e(A) + e(B) - e(AB) \\ \mu_{\ 1-1} &= e(\mathcal{K}) + e(A) - e(B) - e(AB) \\ \mu_{\ 1\ 1} &= e(\mathcal{K}) + e(A) + e(B) + e(AB) \end{aligned} \quad (67)$$

La matrice H des coefficients du système (67) est reportée dans le tableau 5. Ce tableau peut-être utilisé dans les deux sens. Chaque ligne donne les coefficients à appliquer aux effets factoriels pour obtenir l'effet traitement μ_{ij} associé à cette ligne. Chaque colonne donne les coefficients à appliquer aux effets des traitements μ_{ij} pour obtenir, à la division par 4 près, l'effet factoriel associé à cette colonne.

μ_{ij}	$e(\mathcal{K})$	$e(A)$	$e(B)$	$e(AB)$
μ_{-1-1}	1	-1	-1	1
$\mu_{-1\ 1}$	1	-1	1	-1
μ_{1-1}	1	1	-1	-1
$\mu_{1\ 1}$	1	1	1	1

TAB. 5 – *Matrice H de passage des effets factoriels aux effets traitements*

Il est facile de reconstituer ce tableau. Il suffit de remarquer qu'en dehors de la colonne de 1 associée à $e(\mathcal{K})$, on retrouve dans les colonnes $e(A)$, $e(B)$ les niveaux i et j des facteurs A et B et dans la colonne $e(AB)$ les produits ij de ces niveaux. Cette remarque permet d'écrire le système (67) sous la forme plus concise :

$$\mu_{ij} = e(\mathcal{K}) + i e(A) + j e(B) + ij e(AB) \quad (68)$$

Les systèmes (66) et (67) peuvent être aussi réécrits sous la forme matricielle

$$4\mathbf{e} = H'\boldsymbol{\mu} \quad (69)$$

$$\boldsymbol{\mu} = H\mathbf{e} , \quad (70)$$

où $\boldsymbol{\mu}$ et \mathbf{e} sont respectivement le vecteur des effets des traitements et des effets factoriels

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{-1-1} \\ \mu_{-1\ 1} \\ \mu_{1-1} \\ \mu_{1\ 1} \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e(\mathcal{K}) \\ e(A) \\ e(B) \\ e(AB) \end{bmatrix}$$

et H la matrice du tableau 5. Cette matrice vérifie:

$$H'H = 4\mathbf{I} \quad (71)$$

Une telle matrice, formée de 1 et -1 et dont les colonnes sont orthogonales, est appelée *matrice d'Hadamard*.

L'étude des effets factoriels standards définis en (64) répond aux questions qui se posent usuellement avec 2 facteurs. Dans certaines situations cependant, on est amené à définir les effets différemment. Supposons par exemple dans l'exemple des fromages que le type de lait 1 soit en quantité deux fois plus importante que le type de lait -1 dans les mélanges collectés. Il est naturel de faire des moyennes pondérées pour définir l'effet présure:

$$\begin{aligned} \text{moyenne (théorique) de la présure 1} & : \mu_{1\cdot} = \frac{2}{3}\mu_{11} + \frac{1}{3}\mu_{1-1} \\ \text{moyenne (théorique) de la présure -1} & : \mu_{-1\cdot} = \frac{2}{3}\mu_{-11} + \frac{1}{3}\mu_{-1-1} \\ \text{effet pondéré: } p(A) &= \frac{1}{2}(\mu_{1\cdot} - \mu_{-1\cdot}) = \frac{1}{3}(\mu_{11} - \mu_{-11}) + \frac{1}{6}(\mu_{1-1} - \mu_{-1-1}) \end{aligned} \quad (72)$$

La notation avec le point, pour indiquer qu'on fait une moyenne sur l'indice j , est ambiguë, puisque selon le cas il s'agit d'une moyenne non pondérée ou d'une moyenne pondérée. Il est donc essentiel dans chaque cas de bien préciser le système de pondération utilisé.

3.3 Estimation et variance d'estimation

Les estimateurs usuels des paramètres μ_{ij} et σ_{ij}^2 apparaissant dans le modèle (62) sont:

$$\begin{aligned} \hat{\mu}_{ij} &= y_{ij\cdot} \\ \hat{\sigma}_{ij}^2 &= \frac{1}{n_{ij} - 1} \sum_{ijk} (y_{ijk} - y_{ij\cdot})^2 \end{aligned} \quad (73)$$

La variance de $\hat{\mu}_{ij}$ est :

$$\text{var}(\hat{\mu}_{ij}) = \frac{\sigma_{ij}^2}{n_{ij}} \quad (74)$$

On en déduit immédiatement les estimateurs des effets factoriels

$$\begin{aligned} \widehat{e(A)} &= \frac{1}{2}(\hat{\mu}_{1\cdot} - \hat{\mu}_{-1\cdot}) = \frac{1}{4}(-y_{-1-1\cdot} - y_{-11\cdot} + y_{1-1\cdot} + y_{11\cdot}) \\ \widehat{e(B)} &= \frac{1}{2}(\hat{\mu}_{\cdot 1} - \hat{\mu}_{\cdot -1}) = \frac{1}{4}(-y_{-1-1\cdot} + y_{-11\cdot} - y_{1-1\cdot} + y_{11\cdot}) \\ \widehat{e(AB)} &= \frac{1}{4}[(\hat{\mu}_{11} - \hat{\mu}_{-11}) - (\hat{\mu}_{1-1} - \hat{\mu}_{-1-1})] = \frac{1}{4}(+y_{-1-1\cdot} - y_{-11\cdot} - y_{1-1\cdot} + y_{11\cdot}) \end{aligned} \quad (75)$$

et les variances de ces estimateurs qui sont toutes égales à

$$V = \frac{1}{16} \left(\frac{\sigma_{-1-1}^2}{n_{-1-1}} + \frac{\sigma_{-11}^2}{n_{-11}} + \frac{\sigma_{1-1}^2}{n_{1-1}} + \frac{\sigma_{11}^2}{n_{11}} \right) \quad (76)$$

et donc à

$$V = \frac{\sigma^2}{16} \left(\frac{1}{n_{-1-1}} + \frac{1}{n_{-11}} + \frac{1}{n_{1-1}} + \frac{1}{n_{11}} \right) \quad (77)$$

dans le cas homoscédastique où $\sigma_{ij} = \sigma^2$ pour tout traitement ij .

Remarque: On notera que $\hat{\mu}_{i\cdot} = (\hat{\mu}_{i1} + \hat{\mu}_{i-1})/2 = (y_{i1\cdot} + y_{i-1\cdot})/2$ est en général distinct de $y_{i\cdot\cdot}$, puisque :

$$y_{i\cdot\cdot} = \frac{1}{n_{i1} + n_{i-1}} \sum_j \sum_{k=1}^{n_{ij}} y_{ijk} = \frac{n_{i1}y_{i1\cdot} + n_{i-1}y_{i-1\cdot}}{n_{i1} + n_{i-1}}$$

L'estimateur de l'effet pondéré introduit en (72) et la variance de cet estimateur se déduisent également de (73) et (74):

$$\widehat{p(A)} = \frac{1}{3}(y_{11\cdot} - y_{-11\cdot}) + \frac{1}{6}(y_{1-1\cdot} - y_{-1-1\cdot}) \quad (78)$$

$$\text{var}(\widehat{p(A)}) = \frac{1}{9} \frac{\sigma_{11}^2}{n_{11}} + \frac{1}{9} \frac{\sigma_{-11}^2}{n_{-11}} + \frac{1}{36} \frac{\sigma_{1-1}^2}{n_{1-1}} + \frac{1}{36} \frac{\sigma_{-1-1}^2}{n_{-1-1}} \quad (79)$$

3.4 Plan optimal

Les variances des effets factoriels estimés donnés en (75) sont toutes données par la même formule (76). C'est donc le même plan qui minimise ces variances, à savoir celui dans lequel les effectifs sont proportionnels aux quotients des écart-types par les racines carrées des coûts.

$$n_{ij} = \lambda \frac{\sigma_{ij}}{\sqrt{c_{ij}}} \quad (80)$$

En particulier, si les variances et les coûts sont indépendants du traitement, c'est-à-dire si $\sigma_{ij} = \sigma$ et $c_{ij} = 1$ pour tout traitement ij , le plan optimal est celui qui a tous ses effectifs égaux:

$$n_{ij} = \frac{n}{4} . \quad (81)$$

Considérons le cas où l'on s'intéresse à l'effet pondéré $p(A)$ plutôt qu'à l'effet factoriel usuel $e(A)$. Toujours avec la même hypothèse d'égalité des variances et des coûts, le plan qui minimise la variance (79) de $\widehat{p(A)}$ est obtenu lorsque les effectifs pour le lait 1 sont 2 fois ceux pour le lait -1 :

$$n_{i1} = \frac{n}{3} \quad n_{i-1} = \frac{n}{6} \quad (82)$$

Ce n'est donc pas le même plan que celui donné par (81), qui minimise les variances des deux autres effets factoriels estimés $\widehat{e(B)}$ et $\widehat{e(AB)}$. Il est alors naturel de chercher un compromis entre ces deux plans, en minimisant par exemple la somme $\text{var}(\widehat{p(A)}) + \text{var}(\widehat{e(B)})$ des deux variances auxquelles on s'intéresse. En réalité, le déséquilibre entre les effectifs dans le plan (82) est assez réduit. On donne au tableau 6 les variances de $\widehat{p(A)}$ et $\widehat{e(B)}$ dans les plans (81) et (82). Il apparaît que le plan (81) augmente l'écart-type de $\widehat{p(A)}$ de 5% par rapport à celui obtenu dans (82), et qu'inversement le plan (82) augmente l'écart-type de $\widehat{e(B)}$ et $\widehat{e(AB)}$ de 6% par rapport à celui obtenu dans le plan (81).

	Plan (81) $n_{ij} = n/4$	Plan (82) $n_{i1} = n/3 \quad n_{i-1} = n/6$
$\text{var}(\widehat{p(A)})$	$\frac{10}{9} \frac{\sigma^2}{n}$	$\frac{\sigma^2}{n} (\text{optimal})$
$\text{var}(\widehat{e(B)}) = \text{var}(\widehat{e(AB)})$	$\frac{\sigma^2}{n} (\text{optimal})$	$\frac{9}{8} \frac{\sigma^2}{n}$

TAB. 6 – *Variances des effets dans le cas non standard*

Cas où n n'est pas multiple de 4

Si n n'est pas multiple de 4, on ne peut pas trouver des effectifs n_{ij} satisfaisant (81). Dans ce cas, toujours sous l'hypothèse d'égalité des variances et des coûts, les plans optimaux sont ceux qui égalisent le plus possible les effectifs n_{ij} . Ainsi si $n = 14$, les effectifs du plan optimal sont 3,3,4,4. Si $n = 15$ ce sont 3,4,4,4, etc ... Les plans optimaux sont donc ceux pour lesquels la différence entre deux des effectifs n'excède pas 1.

p	0	0.2	0.4	0.6	0.8	1
E	0	0.36	0.64	0.84	0.96	1
$1/\sqrt{E}$	∞	1.67	1.25	1.09	1.02	1

TAB. 7 – *Efficacité avec les effectifs n_1, n_1, n_2, n_2 , où $n_1 = pn/4$*

Pour le montrer, on suppose que la différence entre le plus petit et le plus grand des effectifs est supérieure ou égale à 2. On vérifie alors facilement qu'en augmentant de 1 le plus petit effectif et en diminuant de 1 le plus grand, on diminue strictement la variance V donnée par (77).

3.5 Variance dans les dispositifs déséquilibrés.

On suppose dans ce paragraphe que les variances σ_{ij}^2 sont toutes égales à σ^2 et de même que les coûts c_{ij} sont égaux à 1. On se demande comment varie la variance V des effets factoriels donnée en (77) lorsqu'on s'écarte du plan optimal à effectifs égaux défini par (81). On dit de ce dernier qu'il est *équilibré*, ou *uniforme*. Il s'agit donc de se faire une idée de la perte de précision engendrée par le déséquilibre (la non-uniformité) des effectifs.

On considère deux cas de déséquilibre. Dans le premier, les effectifs sont égaux par paire. Dans le second, trois des effectifs sont égaux et le quatrième est inférieur.

1. Deux des effectifs sont égaux à $pn/4$, les deux autres à $(2-p)n/4$. La variance V est:

$$V = \frac{\sigma^2}{n} \frac{1}{p(2-p)}$$

Le tableau 7 donne les efficacités $E = p(2-p)$ et les coefficients $1/\sqrt{E}$ multiplicateurs de l'écart-type pour quelques valeurs de p . A titre d'exemple, l'écart-type des effets factoriels dans le plan ayant 3,3,7,7 pour effectifs ($p = 0.6$) est de 9% supérieure à celle du plan qui a tous ses effectifs égaux.

On notera que du point de vue variance des effets factoriels standard, les deux plans schématisés ci-dessous sont équivalents:

3	7
3	7

3	7
7	3

Les effectifs du plan de gauche satisfont l'égalité

$$n_{++}n_{ij} = n_{i+}n_{+j} \quad (83)$$

où le + indique une sommation sur l'indice correspondant. On dit parfois d'un tel plan qu'il est *orthogonal au sens large*, la notion d'*orthogonalité au sens strict* étant réservée au plan qui a tous ses effectifs égaux. Ce qui précède montre que, du point de vue variance des effets factoriels, un tel plan n'a aucune vertu spéciale et aucune raison d'être préféré au plan de droite où les effectifs égaux sont en diagonale.

2. L'un des effectifs est égal à $pn/4$, les autres à $\frac{1}{3}(n - pn/4)$. On trouve alors à partir de (77) que:

$$V = \frac{\sigma^2}{n} \frac{1 + 2p}{p(4 - p)}$$

Cette variance est à comparer à la variance minimale σ^2/n . Le rapport

$$E = \frac{\sigma^2/n}{V} = \frac{p(4 - p)}{1 + 2p}$$

est par définition l'efficacité du plan pour l'estimation des effets factoriels. On a reporté dans le tableau 8 les efficacités pour quelques valeurs de p . On a également fait figurer le coefficient par lequel est multiplié l'écart-type des effets factoriels, qui est $1/\sqrt{E}$.

p	0	0.1	0.2	0.3	0.4	0.5
E	0	0.33	0.54	0.69	0.80	0.88
$1/\sqrt{E}$	∞	1.76	1.36	1.20	1.12	1.07

TAB. 8 – *Efficacité avec trois des effectifs égaux et le quatrième égal à $pn/4$*

A titre d'exemple, le plan qui a pour effectif 3,7,7,7 ($p = 0.5$) donne aux effets factoriels estimés un écart-type de 7% supérieur à celui du plan 6,6,6,6.

Dans le cas $p = 0$ où l'un des effectifs est nul, disons $n_{11} = 0$, la moyenne correspondante μ_{11} est *non estimable* et par suite aucun des effets factoriels n'est estimable. Cela ne signifie évidemment pas que les résultats d'un tel plan ne présente aucun intérêt. On peut toujours comparer les moyennes μ_{ij} des cases non vides pour lesquelles $n_{ij} \neq 0$. D'autre part,

nous allons voir que sous l'hypothèse de nullité de l'interaction $e(AB)$, les effets principaux $e(A)$ et $e(B)$ deviennent estimables. Les estimations correspondantes, sans biais si $e(AB) = 0$, sont lorsque $e(AB) \neq 0$ biaisées par une quantité qui dépend du terme $e(AB)$ négligé.

Si p n'est pas nul, mais qu'il est très inférieur à 1, tous les effets factoriels sont estimables dans le cadre du modèle (62). Cependant les estimateurs biaisés de $e(A)$ et $e(B)$ obtenus en faisant l'hypothèse de nullité de $e(AB)$ peuvent être meilleurs du point de vue écart quadratique moyen que ceux donnés en (75). Nous allons donc également étudier dans quelles conditions on a avantage à utiliser ces estimateurs biaisés.

3.6 Estimation dans le cadre du modèle additif

Nous considérons à titre d'exemple le plan défini par les effectifs:

		$B(j)$	
		-1	1
$A(i)$	-1	0	1
	1	1	2

Prenant en compte la décomposition (68) de μ_{ij} en fonction des effets factoriels, on peut réécrire le modèle $y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, où $\text{var}(\varepsilon_{ijk}) = \sigma^2$, sous la forme

$$y_{ijk} = e(\mathcal{K}) + i e(A) + j e(B) + ij e(AB) + \varepsilon_{ijk} \quad \text{var}(\varepsilon_{ijk}) = \sigma^2 \quad (84)$$

Appliqué aux 4 observations, ce modèle donne les 4 equations

$$\begin{aligned} y_{1-11} &= e(\mathcal{K}) + e(A) - e(B) - e(AB) + \varepsilon_{1-11} \\ y_{111} &= e(\mathcal{K}) + e(A) + e(B) + e(AB) + \varepsilon_{111} \\ y_{112} &= e(\mathcal{K}) + e(A) + e(B) + e(AB) + \varepsilon_{112} \\ y_{-111} &= e(\mathcal{K}) - e(A) + e(B) - e(AB) + \varepsilon_{-111} \end{aligned}$$

qui s'écrivent sous forme matricielle

$$\begin{bmatrix} y_{1-11} \\ y_{111} \\ y_{112} \\ y_{-111} \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \times \begin{bmatrix} e(\mathbb{K}) \\ e(A) \\ e(B) \\ e(AB) \end{bmatrix} + \begin{bmatrix} \varepsilon_{1-11} \\ \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{-111} \end{bmatrix}$$

$$\mathbf{y} = X \mathbf{e} + \boldsymbol{\varepsilon}$$

A cause de la case vide, on ne peut pas estimer les 4 effets factoriels du vecteur \mathbf{e} , ce qui se traduit par le fait que $X'X$ n'est pas inversible. Mais si l'on suppose l'interaction $e(AB)$ nulle, on peut supprimer cet effet dans \mathbf{e} , la colonne correspondante dans X et trouver alors $\widehat{e(A)}$ et $\widehat{e(B)}$ par les formules usuelles. Le modèle (84) sans le terme $e(AB)$ d'interaction sera appelé dans ce qui suit *modèle additif*.

Pour simplifier les calculs, on opère une "reparamétrisation" du modèle permettant de centrer les colonnes de X autres que la première. Les moyennes des colonnes 2, 3, 4 étant 0.5, 0.5, 0, on écrit le modèle (84) sous la forme:

$$y_{ijk} = \mu + (i - 0.5)e(A) + (j - 0.5)e(B) + ije(AB) + \varepsilon_{ijk} \quad (85)$$

où le seul nouveau paramètre μ est :

$$\mu = e(\mathbb{K}) + 0.5e(A) + 0.5e(B)$$

Ce modèle modifié conduit au système matriciel:

$$\underbrace{\begin{bmatrix} y_{1-11} \\ y_{111} \\ y_{112} \\ y_{-111} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & 0.5 & -1.5 \\ 1 & 0.5 & 0.5 \\ 1 & 0.5 & 0.5 \\ 1 & -1.5 & 0.5 \end{bmatrix}}_{X_1} \underbrace{\begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix}}_{X_2} \times \begin{bmatrix} \mu \\ e(A) \\ \frac{e(B)}{e(AB)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1-11} \\ \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{-111} \end{bmatrix}$$

$$\mathbf{y} = [X_1 \quad X_2] \times \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} + \boldsymbol{\varepsilon}$$

qu'on peut encore réécrire sous la forme :

$$\mathbf{y} = X_1 \mathbf{e}_1 + X_2 \mathbf{e}_2 + \boldsymbol{\varepsilon} \quad (86)$$

où X_1 et X_2 sont les deux matrices indiquées ci-dessus et

$$\mathbf{e}_1 = \begin{bmatrix} \mu \\ e(A) \\ e(B) \end{bmatrix} \quad \mathbf{e}_2 = [e(AB)]$$

Faisons provisoirement l'hypothèse que le modèle est additif; c'est-à-dire que :

$$e(AB) = 0 \quad (87)$$

Le modèle (86) devient

$$\mathbf{y} = X_1 \mathbf{e}_1 + \boldsymbol{\varepsilon}$$

L'estimation des moindres carrés de \mathbf{e}_1 et sa matrice de covariances sont donnés par :

$$\hat{\mathbf{e}}_1 = (X_1' X_1)^{-1} X_1' \mathbf{y} \quad \text{cov}(\hat{\mathbf{e}}_1) = \sigma^2 (X_1' X_1)^{-1} \quad (88)$$

Le calcul donne :

$$X_1' X_1 = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & -1 \\ 0 & -1 & 3 \end{bmatrix} \quad (X_1' X_1)^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 3/8 & 1/8 \\ 0 & 1/8 & 3/8 \end{bmatrix}$$

$$(X_1' X_1)^{-1} X_1' = \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & -2 \\ -2 & 1 & 1 & 0 \end{bmatrix}$$

$$\hat{\mathbf{e}}_1 = \begin{bmatrix} \hat{\mu} \\ \widehat{e(A)} \\ \widehat{e(B)} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} y_{1-11} & + & y_{111} & + & y_{112} & + & y_{-111} \\ -2y_{1-11} & + & y_{111} & + & y_{112} & - & 2y_{-111} \end{bmatrix} = \begin{bmatrix} y_{\dots} \\ \frac{1}{2}(y_{11\cdot} - y_{-11\cdot}) \\ \frac{1}{2}(y_{11\cdot} - y_{1-1\cdot}) \end{bmatrix}.$$

L'effet principal de A est donc estimé à partir des seuls résultats relatifs au niveau 1 de B:

$$\widehat{e(A)} = \frac{1}{2}(y_{11\cdot} - y_{-11\cdot}).$$

De même, l'effet principal de B est estimé à partir des seuls résultats relatifs au niveau 1 de A:

$$\widehat{e(B)} = \frac{1}{2}(y_{11\cdot} - y_{1-1\cdot}).$$

Si l'interaction est effectivement nulle, c'est à dire si $e(AB) = 0$, ces estimateurs sont non biaisés. Il est cependant intéressant d'étudier par quelle

quantité ils sont biaisés lorsque $e(AB) \neq 0$. On examine pour cela l'espérance de $\hat{\mathbf{e}}_1$:

$$E(\hat{\mathbf{e}}_1) = \mathbf{e}_1 + (X_1' X_1)^{-1} X_1' X_2 \mathbf{e}_2 \quad (89)$$

Le *biais d'estimation* de \mathbf{e}_1 est donc $(X_1' X_1)^{-1} X_1' X_2 \mathbf{e}_2$. On a:

$$X_1' X_2 = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} \quad (X_1' X_1)^{-1} X_1' X_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

d'où

$$E \begin{pmatrix} \hat{\mu} \\ \widehat{e(A)} \\ \widehat{e(B)} \end{pmatrix} = \begin{pmatrix} \mu \\ e(A) \\ e(B) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} e(AB)$$

soit

$$E(\widehat{e(A)}) = e(A) + e(AB)$$

$$E(\widehat{e(B)}) = e(B) + e(AB)$$

Autrement dit, si l'interaction $e(AB)$ n'est pas nulle, l'estimateur $\widehat{e(A)}$ est centré, non pas sur l'effet $e(A)$ que l'on cherche à estimer, mais sur $e(A) + e(AB)$. Si l'on pouvait répéter la même expérience un très grand nombre de fois, la série des estimations obtenues serait centrée non pas sur l'effet $e(A)$, mais sur $E(\widehat{e(A)}) = e(A) + e(AB)$, comme cela est schématisé sur la figure 1.

L'erreur d'estimation $\widehat{e(A)} - e(A)$ se décompose donc sous la forme d'une erreur non systématique $\widehat{e(A)} - E(\widehat{e(A)})$ et d'une erreur systématique ou biais $E(\widehat{e(A)}) - e(A) = e(AB)$. La qualité de l'estimateur $\widehat{e(A)}$ est mesuré par l'espérance du carré de cette erreur, qui se décompose en somme de la variance et du carré du biais:

$$E \left(\left(\widehat{e(A)} - e(A) \right)^2 \right) = \text{var} \left(\widehat{e(A)} \right) + \left(E \left(\widehat{e(A)} \right) - e(A) \right)^2$$

Cette espérance est appelée *Ecart Quadratique Moyen* (Mean Square Error), ce qui s'abrège en EQM (en anglais MSE). Elle mesure ce que serait la moyenne du carré de l'erreur sur une longue série d'expériences analogues.

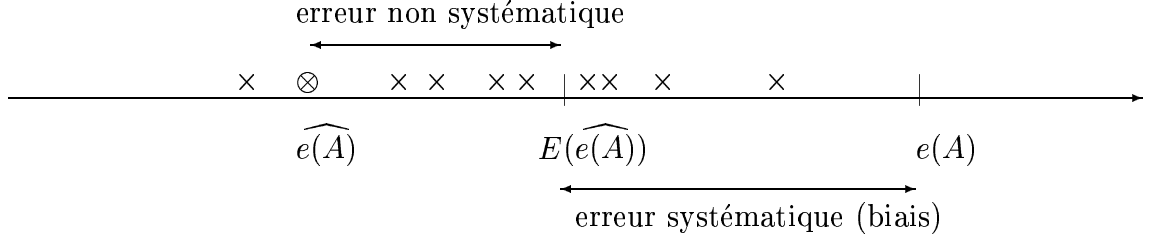


FIG. 1 – *Décomposition de l'erreur dans le cas d'une estimation biaisée*

Les \times symbolisent les estimations qui seraient obtenues en répétant exactement la même expérience. Le \otimes symbolise l'estimation obtenue lors de l'expérience réellement effectuée.

La variance de $e(\widehat{A})$ est $\frac{3}{8}\sigma^2$ et le biais $e(AB)$. L'écart quadratique moyen est donc

$$\text{EQM}(e(\widehat{A})) = \frac{3}{8}\sigma^2 + e(AB)^2 .$$

Cette mesure EQM de qualité de l'estimateur n'a pas d'intérêt dans cet exemple, puisque on est obligé, si l'on veut estimer $e(A)$ et $e(B)$, de supposer la nullité de $e(AB)$. Elle devient intéressante dans un cas limite tel que celui du plan du tableau 9, où l'effectif m est nettement supérieur à 1. Si on prend

1	m
m	m

TAB. 9 – *Un plan fortement déséquilibré*

$m = 13$ par exemple, le tableau 8 montre que l'efficacité est de $E = 0.33$ (prendre $n = 40, p = 0.1$). La variance est donc près de 3 fois plus grande que celle du plan ayant 10 répétitions par case, et les estimateurs biaisés du modèle additif peuvent être préférable aux estimateurs sans biais à forte variance donnés en (75).

Les calculs donnant ces estimateurs du modèle additif, leurs variances et biais d'estimation sont les mêmes que précédemment. Les moyennes des

niveaux de A , B et AB sont a , a et $-a$ respectivement, où a est donné par :

$$a = \frac{m-1}{n}, \quad n = 3m+1$$

Au modèle reparamétrisé:

$$y_{ijk} = \mu + (i-a)e(A) + (j-a)e(B) + (ij+a)e(AB) + \varepsilon_{ijk}$$

correspond les matrices X_1 et X_2 schématisée ci-dessous:

$\not\sim$	\tilde{A}	\tilde{B}	\widetilde{AB}	
1	$-1-a$	$-1-a$	$1+a$	
1	$1-a$	$-1-a$	$-1+a$	$\updownarrow m$ lignes
\dots	\dots	\dots	\dots	
1	$-1-a$	$1-a$	$-1+a$	$\updownarrow m$ lignes
\dots	\dots	\dots	\dots	
1	$1-a$	$1-a$	$1+a$	$\updownarrow m$ lignes
\dots	\dots	\dots	\dots	

X_1
 X_2

Les formules matricielles (88) et (89) permettent, après un calcul un peu laborieux, d'obtenir les estimations de $e(A)$, $e(B)$, ainsi que les variances et biais d'estimation. Compte tenu de la symétrie entre A et B , nous ne donnons que les formules relatives à $e(A)$:

$$\widehat{e(A)} = \frac{1}{m+3} \left[\frac{m+1}{2} (y_{11\cdot} - y_{-11\cdot}) + (y_{1-1\cdot} - y_{-1-1\cdot}) \right] \quad (90)$$

$$\text{var} \left(\widehat{e(A)} \right) = \frac{m+1}{2m(m+3)} \sigma^2 \quad \text{Biais} = \frac{m-1}{m+3} e(AB) \quad (91)$$

On remarque que $\widehat{e(A)}$ est, à une division par 2 près, une moyenne pondérée des effets estimés de A aux deux niveaux de B . Les poids sont inversement proportionnels aux variances de ces effets estimés:

$$\text{var}(y_{11\cdot} - y_{-11\cdot}) = 2 \frac{\sigma^2}{m} \quad \text{var}(y_{1-1\cdot} - y_{-1-1\cdot}) = \left(1 + \frac{1}{m} \right) \sigma^2$$

On retrouve facilement les expressions (91) de la variance et du biais à partir de l'expression (90) de l'estimateur du modèle additif.

L'écart quadratique moyen est:

$$\text{EQM}(\widehat{e(A)}) = \sigma^2 \left[\frac{m+1}{2m(m+3)} + \frac{(m-1)^2}{(m+3)^2} \left(\frac{e(AB)}{\sigma} \right)^2 \right]$$

Il peut être estimé en remplaçant σ et $e(AB)$ par leurs estimations:

$$\widehat{\text{EQM}}(\widehat{e(A)}) = \hat{\sigma}^2 \left[\frac{m+1}{2m(m+3)} + \frac{(m-1)^2}{(m+3)^2} \left(\frac{\widehat{e(AB)}}{\hat{\sigma}} \right)^2 \right]$$

Cet écart quadratique est à comparer à celui qu'on obtient en estimant $e(A)$ dans le cadre du modèle avec interaction. On note ici $\widetilde{e(A)}$ l'estimateur correspondant, pour le distinguer de $\widehat{e(A)}$. Rappelons (75 et 77) que:

$$\widetilde{e(A)} = \frac{1}{4} [(y_{11\cdot} - y_{-11\cdot}) + (y_{1-1\cdot} - y_{-1-1\cdot})]$$

$$\text{var}(\widetilde{e(A)}) = \frac{\sigma^2}{16} \left(1 + \frac{3}{m} \right) \quad \text{Biais}(\widetilde{e(A)}) = 0$$

$$\text{EQM}(\widetilde{e(A)}) = \text{var}(\widetilde{e(A)}) = \frac{\sigma^2}{16} \left(1 + \frac{3}{m} \right)$$

Le ratio des EQM des deux estimateurs est donc:

$$\begin{aligned} \frac{\text{EQM}(\widehat{e(A)})}{\text{EQM}(\widetilde{e(A)})} &= \frac{8(m+1)}{(m+3)^2} + \frac{16m(m-1)^2}{(m+3)^3} \left(\frac{e(AB)}{\sigma} \right)^2 \\ &= v + b \left(\frac{e(AB)}{\sigma} \right)^2 \end{aligned}$$

avec:

$$v = \frac{8(m+1)}{(m+3)^2} \quad b = \frac{16m(m-1)^2}{(m+3)^3}$$

Le tableau 10 donne les valeurs de v et b pour des valeurs de m allant de 1 à 10, ainsi que la valeur maximum que doit avoir le rapport $R = (e(AB)/\sigma)^2$

m	1	2	3	4	5	6	7	8	9	10
v	1.00	0.96	0.89	0.82	0.75	0.69	0.64	0.60	0.56	0.52
b	0.00	0.26	0.89	1.68	2.50	3.29	4.03	4.71	5.33	5.90
R	—	0.16	0.13	0.11	0.10	0.09	0.09	0.09	0.08	0.08

TAB. 10 – *Comparaison des estimateurs des modèles avec et sans interaction*

pour que l'estimateur du modèle additif soit meilleur que celui du modèle interactif.

La variance de $\widehat{e(AB)}$ est $\frac{\sigma^2}{16} (1 + \frac{3}{m})$. L'interaction sera jugée significativement différente de 0 si

$$\frac{|\widehat{e(AB)}|}{\hat{\sigma}} \geq t_\alpha \frac{1}{4} \sqrt{1 + \frac{3}{m}} \quad (92)$$

où t_α est le seuil ad hoc de la loi de Student. Prenons $\alpha = 5\%$. On a alors $t_\alpha \simeq 2$ si le nombre de degrés de liberté $3(m-1)$ est assez grand et $t_\alpha \geq 2$ dans les autres cas. La valeur à droite de (92) est minorée par $1/2$ et il faut que $\hat{R} = \left(\widehat{e(AB)}/\hat{\sigma}\right)^2$ soit supérieur à 0.25 pour que $e(AB)$ soit significativement différent de 0 au seuil $\alpha = 5\%$. Le tableau 10 montre alors que quel que soit le degré de déséquilibre, il y a un large domaine de variation de $e(AB)/\sigma$ pour lequel l'estimateur du modèle avec interaction est meilleur que l'estimateur du modèle additif, bien qu'on ait fort peu de chance de décréter l'interaction significative. Notons aussi que même si $\widehat{e(AB)}$ est nul, il n'est pas du tout invraisemblable que $e(AB)/\sigma$ ait une valeur qui rende l'estimateur du modèle avec interaction meilleur que celui du modèle additif.

En conclusion, dans l'analyse d'un dispositif déséquilibré tel que celui du tableau 9, même si l'interaction est non significativement différente de 0, il peut être préférable d'estimer les effets principaux $e(A)$ et $e(B)$ dans le modèle avec interaction plutôt que dans le modèle additif. Si cette estimation sous le modèle avec interaction conduit à des effets non "significatifs", alors que le modèle additif conduit à des effets significatifs, il faut se demander quel est le sens exact des différences trouvées. On examine éventuellement des différences autres que les effets factoriels (par exemple $\mu_{11} - \mu_{-11}$) pour arriver à des conclusions nuancées telle que: il y a un effet de A pour le niveau 1 de B , mais on ne peut pas conclure à cause du manque d'information pour le niveau -1 de B .

Il est évidemment toujours préférable quand on s'intéresse à des effets

standards d'essayer d'avoir un plan aussi équilibré que possible. On minimise ainsi la variance, et nous allons voir qu'on minimise aussi le biais lorsqu'on utilise un modèle additif pour l'estimation de $e(A)$ et $e(B)$.

3.7 Plans optimaux pour la minimisation du biais

Plaçons nous dans le cas où l'on adopte systématiquement les estimations des moindres carrés du modèle additif pour $e(A)$ et $e(B)$. Le biais est alors donné par la quantité $(X_1'X_1)^{-1}X_1'X_2\mathbf{e}_2$ qui apparaît dans (89). Il est nul, et le carré du biais est donc minimal, lorsque $X_1'X_2 = 0$. Cette égalité est en particulier vérifiée si tous les effectifs n_{ij} sont égaux à $n/4$. Le plan à effectifs égaux, optimal du point de vue variance, minimise donc également le biais si on utilise un modèle additif pour estimer alors que le vrai modèle comporte une interaction.

En fait, il existe d'autres plans pour lesquels $X_1'X_2$ est nul et le biais dû à l'interaction nul lorsqu'on estime sous le modèle additif. Ce sont les plans du type schématisé ci-dessous:

m_1	0	0	m_1
0	m_2	m_2	0

Mais ces plans sont clairement à rejeter, car ils rendent les effets principaux $e(A)$ et $e(B)$ non estimables, même dans le cadre du modèle additif. Ainsi le premier de ces plans permet d'estimer μ_{-1-1} et μ_{11} . Or sous l'hypothèse d'additivité, on a

$$\begin{aligned}\mu_{-1-1} &= e(\mathbb{K}) - e(A) - e(B) \\ \mu_{11} &= e(\mathbb{K}) + e(A) + e(B)\end{aligned}$$

et ces équations ne permettent pas d'exprimer $e(A)$ et $e(B)$ en fonction de μ_{-1-1} et μ_{11} et $e(A)$.

4 Régression simple

4.1 Modèle et estimation

On suppose que la réponse y dépend linéairement d'une variable quantitative x sur un intervalle borné. Quitte à effectuer un changement d'origine

et d'échelle sur x , on peut écrire le modèle pour $x \in [-1,1]$:

$$y = \alpha + \beta x + \varepsilon \quad E(\varepsilon) = 0 . \quad (93)$$

La variance de ε est supposé constante:

$$\text{var}(\varepsilon) = \sigma^2 . \quad (94)$$

Le nombre n d'observations étant fixé, on cherche les valeurs x_1, \dots, x_n à donner à x pour estimer le plus précisément possible soit la pente β , soit les réponses $f(x) = \alpha + \beta x$ pour x variant dans $[-1,1]$. Avant d'aborder ce problème de planification, rappelons comment on estime β et $f(x)$ lorsqu'on connaît x_1, \dots, x_n .

Estimation et variance d'estimation

Il est commode de reparamétriser le modèle en centrant la variable explicative x sur la moyenne $x_{\bullet} = \sum x_i/n$:

$$y = \mu + \beta(x - x_{\bullet}) + \varepsilon \quad \text{où } \mu = \alpha + \beta x_{\bullet} . \quad (95)$$

Appliqué aux n valeurs x_1, \dots, x_n , ce modèle conduit au système "matriciel":

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 - x_{\bullet} \\ \vdots & \vdots \\ 1 & x_n - x_{\bullet} \end{bmatrix} \times \begin{bmatrix} \mu \\ \beta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

que l'on écrit sous la forme

$$\mathbf{y} = X\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

avec

$$\boldsymbol{\theta} = \begin{bmatrix} \mu \\ \beta \end{bmatrix}$$

On a

$$X'X = n \begin{bmatrix} 1 & 0 \\ 0 & m_2 \end{bmatrix} \quad X'Y = \begin{bmatrix} \sum y_i \\ \sum_i (x_i - x_{\bullet}) y_i \end{bmatrix}$$

où

$$m_2 = \sum_i (x_i - x_{\bullet})^2 / n \quad (96)$$

Les estimateurs de μ, β sont donc

$$\hat{\mu} = \frac{\sum y_i}{n} \quad \hat{\beta} = \frac{\sum (x_i - x_{\bullet}) y_i}{nm_2} \quad (97)$$

Ils ont

$$\text{cov} \begin{pmatrix} \hat{\mu} \\ \hat{\beta} \end{pmatrix} = \sigma^2 (X'X)^{-1} = \frac{\sigma^2}{n} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{m_2} \end{bmatrix}. \quad (98)$$

pour matrice de covariance et on a donc

$$\text{var}(\hat{\mu}) = \frac{\sigma^2}{n} \quad \text{var}(\hat{\beta}) = \frac{\sigma^2}{n} \frac{1}{m_2} \quad \text{cov}(\hat{\mu}, \hat{\beta}) = 0 \quad (99)$$

On en déduit l'estimation de $f(x)$ et la variance correspondante

$$\widehat{f(x)} = \hat{\mu} + \hat{\beta}(x - x_{\bullet}) \quad (100)$$

$$\text{var} \left(\widehat{f(x)} \right) = \frac{\sigma^2}{n} \left[1 + \frac{(x - x_{\bullet})^2}{m_2} \right] \quad (101)$$

4.2 Plan optimal

4.2.1 Plan optimal pour l'estimation de la pente

La variance de $\hat{\beta}$ donné en (99) est minimum lorsque $m_2 = \sum (x_i - x_{\bullet})^2 / n$ est maximum.

Proposition 4.1 *Les plans qui minimisent la variance de $\hat{\beta}$ sont ceux qui répartissent équitablement les x_i sur les seules extrémités $-1, 1$ du domaine de définition de x .*

En d'autres termes, si n est pair, l'optimum est obtenu lorsque la moitié des x_i sont égaux à 1, l'autre moitié à -1 . Si $n = 2m + 1$ est impair, on place m valeurs à une extrémité (1 ou -1) et $m + 1$ valeurs à l'autre extrémité. Il y a donc dans ce cas deux plans optimaux. Lorsque n est pair, le plan optimum donne $m_2 = 1$, soit:

$$\min \text{var}(\hat{\beta}) = \frac{\sigma^2}{n}$$

Démonstration de la proposition 4.1

Il s'agit de rendre maximum $v = \sum_i (x_i - x_{\bullet})^2$. Si on note δ_{ij} le symbole de Kronecker, on a

$$\frac{\partial v}{\partial x_j} = \sum_i 2(x_i - x_{\bullet})(\delta_{ij} - \frac{1}{n}) = 2(x_j - x_{\bullet}).$$

Si $x_{\bullet} < x_j < 1$, cette dérivée partielle est strictement positive et une légère augmentation de x_j provoque une augmentation de v . De même si $-1 < x_j < x_{\bullet}$, on augmente v en diminuant x_j . D'autre part si $x_j = x_{\bullet}$, on a $\frac{\partial v}{\partial x_j} = 0$ et $\frac{\partial^2 v}{\partial x_j^2} = 2(1 - \frac{1}{n}) > 0$. Le développement de Taylor de v en fonction de x_j

$$v(x_j) = v(x_{\bullet}) + \left(1 - \frac{1}{n} + o(x_j - x_{\bullet})\right) (x_j - x_{\bullet})^2,$$

où $o(x_j - x_{\bullet})$ tend vers 0 avec $(x_j - x_{\bullet})$, montre alors qu'on peut augmenter v en modifiant légèrement x_j .

Il résulte de ce qui précède qu'au maximum de v , les x_i sont tous égaux à 1 ou -1 . Si m est le nombre de ceux qui sont égaux à -1 , on a $x_{\bullet} = n - 2m/n$ et

$$v = (n - m) \left(1 - \frac{n - 2m}{n}\right)^2 + m \left(-1 - \frac{n - 2m}{n}\right)^2 = \frac{4m(n - m)}{n}$$

Il est alors facile de montrer que v est maximum lorsque m et $n - m$ diffèrent au plus de 1 \square

4.2.2 Plan optimal pour l'estimation de la réponse théorique

Pour x fixé, la variance de $\widehat{f(x)}$ donnée en (101) est minimum si et seulement si le plan vérifie $x_{\bullet} = x$. On ne peut donc pas trouver un unique plan qui rende minimum la variance de $\widehat{f(x)}$ quel que soit $x \in [-1, 1]$ et il faut trouver un compromis. On présente ici deux façons de définir un tel compromis.

4.2.2.1 Critère G et théorème de Kiefer-Wolfowitz

La qualité globale des prédictions est mesurée par la plus grande des valeurs de $\text{var}(\widehat{f(x)})$ pour $x \in [-1, 1]$:

$$\max_x \frac{\sigma^2}{n} \left(1 + \frac{(x - x_{\bullet})^2}{m_2}\right).$$

On minimise cette quantité, ou de façon équivalente le produit par n/σ^2 de cette quantité qui est appelé le *critère* G :

$$G = \max_x \left(1 + \frac{(x - x_{\cdot})^2}{m_2} \right)$$

Proposition 4.2 *Si n est pair, les assertions suivantes sont équivalentes:*

- i) le plan minimise la variance de $\hat{\beta}$*
- ii) le plan minimise G*
- iii) $G = 2$ (nombre de paramètres de l'espérance dans le modèle (93)).*

Cette proposition est un cas très particulier du théorème de Kiefer-Wolfowitz (1960). L'assertion *i*) est usuellement écrite sous la forme:

i') le plan minimise $\det[n(X'X)^{-1}]$

D'après (98), ce déterminant est égal à $1/m_2 = \sigma^2 \text{var}(\hat{\beta})/n$ et *i*) et *i')* sont donc équivalents.

On vérifie ici immédiatement que le plan qui répartit les x_i pour moitié en 1, pour moitié en -1 vérifie l'assertion *iii*). La proposition 4.2 (i.e. le théorème de Kiefer Wolfowitz) permet alors d'affirmer que ce plan minimise G et $\text{var}(\hat{\beta})$. La proposition 4.1 apparaît donc, lorsque n est pair, comme un corollaire.

La démonstration de la proposition 4.2 sort du cadre de ce cours, mais il est par contre facile de montrer directement que le plan qui minimise G est, dans le cas n pair, le même que celui qui minimise $\text{var}(\hat{\beta})$. On a en effet

$$\max_x (x - x_{\cdot})^2 = \max \left((1 - x_{\cdot})^2, (-1 - x_{\cdot})^2 \right) \geq 1$$

d'où

$$G \geq 1 + \frac{1}{m_2} \geq 1 + \frac{1}{\max m_2} = 2$$

Comme $G = 2$ pour le plan qui place la moitié des x_i en 1, l'autre moitié en -1 , ce plan est G optimal.

4.2.2.2 Critère de variance moyenne

La qualité globale de la prédiction est mesurée par la moyenne sur $x \in [-1,1]$ de $\text{var}(\widehat{f(x)})$. On minimise cette quantité, ou de façon équivalente le produit V par n/σ^2 de cette quantité:

$$V = \frac{1}{2} \int_{-1}^{+1} \left[1 + \frac{(x - x_{\cdot})^2}{m_2} \right] dx = 1 + \frac{1 + 3x_{\cdot}^2}{3m_2} . \quad (102)$$

Ce critère est minimum lorsque m_2 est maximum. Sa minimisation conduit donc au même plan optimal que celui donné dans la proposition 4.1.

4.3 Prise en compte du biais

La solution optimale trouvée précédemment peut s'avérer mauvaise si la variation de y en fonction de x n'est plus strictement linéaire. Pour préciser cela, supposons que la variation de y en fonction de x sur $[-1,1]$ ne soit plus strictement linéaire mais de la forme

$$y = f(x) + \varepsilon \quad (103)$$

avec

$$f(x) = \beta_0 + \beta_1 x + \beta_{11} x^2 \quad (104)$$

et

$$E(\varepsilon) = 0 , \quad \text{var}(\varepsilon) = \sigma^2 .$$

Le plan qui distribue équitablement les points sur $-1, +1$ a deux gros défauts.

- Il ne permet pas de détecter la non linéarité de la relation entre y et x , c'est-à-dire l'existence d'un coefficient β_{11} non nul.
- La droite de regression, qui joint les points moyens associés aux valeurs 1 et -1 de x , donne une erreur systématique très importante au centre de l'intervalle $[-1,1]$.

Nous allons examiner, en suivant le développement de Box et Draper dans *A basis for the selection of a Response surface Design, 1959* (Collected Works of Box, p 456), comment l'estimateur de $f(x)$ obtenu sous l'hypothèse de linéarité (93) est biaisé par la présence d'un terme quadratique $\beta_{11}x^2$ dans la réponse théorique, et comment on peut prendre ce biais en compte dans la recherche d'un bon plan d'expérience.

Après centrage de x , la réponse théorique $f(x)$ donnée en (104) devient

$$f(x) = \mu + \beta(x - x_{\bullet}) + \beta_{11}(x - x_{\bullet})^2 \quad (105)$$

avec

$$\mu = \beta_0 + \beta_1 x_{\bullet} + \beta_{11} x_{\bullet}^2$$

$$\beta = \beta_1 + 2\beta_{11} x_{\bullet}$$

Appliqué à x_1, \dots, x_n , le modèle (103) donne alors

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 - x_{\bullet} \\ \vdots & \vdots \\ 1 & x_n - x_{\bullet} \end{bmatrix} \times \begin{bmatrix} \mu \\ \beta \end{bmatrix} + \begin{bmatrix} (x_1 - x_{\bullet})^2 \\ \vdots \\ (x_n - x_{\bullet})^2 \end{bmatrix} \beta_{11} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{y} = X_1 \boldsymbol{\theta} + X_2 \beta_{11} + \boldsymbol{\varepsilon}$$

Le vecteur $\boldsymbol{\theta}$ est estimé, sous l'hypothèse $\beta_{11} = 0$, par:

$$\hat{\boldsymbol{\theta}} = (X_1' X_1)^{-1} X_1' \mathbf{y}$$

Il a pour espérance;

$$E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta} + (X_1' X_1)^{-1} X_1' X_2 \beta_{11}$$

Or

$$(X_1' X_1)^{-1} X_1' X_2 = \begin{bmatrix} m_2 \\ \frac{m_3}{m_2} \end{bmatrix}$$

où m_3 est défini de façon analogue à m_2 en (96):

$$m_3 = \sum_i (x_i - x_{\bullet})^3 / n \quad (106)$$

On a donc:

$$E(\hat{\mu}) = \mu + m_2 \beta_{11}$$

$$E(\hat{\beta}) = \beta + \frac{m_3}{m_2} \beta_{11} .$$

L'estimateur des moindres carrés de $f(x)$ sous l'hypothèse $\beta_{11} = 0$ est alors donnée par l'expression (100). Sa variance est donnée par (101). Son espérance (lorsque β_{11} n'est plus supposé nul) est

$$E(\hat{\mu} + \hat{\beta}(x - x_{\bullet})) = \mu + \beta(x - x_{\bullet}) + m_2\beta_{11} + \frac{m_3}{m_2}\beta_{11}(x - x_{\bullet}) .$$

Le biais, différence entre cette espérance et la réponse donnée par (105), est donc

$$m_2\beta_{11} + \frac{m_3}{m_2}\beta_{11}(x - x_{\bullet}) - \beta_{11}(x - x_{\bullet})^2 .$$

Le carré de ce biais multiplié par n/σ^2 est noté $B(x)$:

$$B(x) = \alpha_{11}^2 \left(m_2 - (x - x_{\bullet})^2 + \frac{m_3}{m_2}(x - x_{\bullet})^2 \right) \quad (107)$$

où

$$\alpha_{11}^2 = \frac{n}{\sigma^2} \beta_{11}^2 . \quad (108)$$

On note de même $V(x)$ et $J(x)$ le produit par n/σ^2 de la variance et de l'écart quadratique de $\widehat{f(x)}$:

$$V(x) = 1 + \frac{(x - x_{\bullet})^2}{m_2} \quad (109)$$

$$J(x) = V(x) + B(x) \quad (110)$$

Il est clair qu'on ne peut pas minimiser $J(x)$, donc l'écart quadratique, simultanément pour tous les x . On cherche donc un compromis en minimisant par exemple la moyenne J de $J(x)$ sur $[-1,1]$ qui est:

$$J = \frac{1}{2} \int_{-1}^1 J(x) dx = V + B \quad (111)$$

où V est la moyenne de $V(x)$ donnée en (102) et B de même la moyenne du biais $B(x)$. Pour simplifier les calculs, nous nous plaçons dans le cas où l'on impose *a priori* aux x_i d'être centré sur 0:

$$x_{\bullet} = 0. \quad (112)$$

Dans ce cas l'intégration de $B(x)$ donne

$$B = \frac{1}{2} \int_{-1}^1 B(x) dx = \alpha_{11}^2 \left[m_2^2 - \frac{2m_2}{3} + \frac{1}{5} + \frac{m_3^2}{3m_2^2} \right] \quad (113)$$

et on a donc

$$J = \left[1 + \frac{1}{3m_2}\right] + \alpha_{11}^2 \left[m_2^2 - \frac{2m_2}{3} + \frac{1}{5} + \frac{m_3^2}{3m_2^2}\right]. \quad (114)$$

Pour minimiser J , on a clairement intérêt à annuler m_3 en prenant par exemple une distribution symétrique. On a alors :

$$J = \left[1 + \frac{1}{3m_2}\right] + \alpha_{11}^2 \left[\left(m_2 - \frac{1}{3}\right)^2 + \frac{4}{45}\right] \quad (115)$$

La partie biais est minimum quand $m_2 = 1/3$, ce qui donne une valeur moyenne de $1/3$ pour x_i^2 soit une valeur de $1/\sqrt{3} = 0.58$ pour $|x_i|$.

La minimisation de J conduit alors à sélectionner pour m_2 une valeur intermédiaire entre 0.58 et 1, en fonction de l'importance de α_{11} , qui est le rapport entre l'effet quadratique β_{11} et le coefficient σ/\sqrt{n} .

On voit donc que la prise en compte d'un éventuel biais causé par un écart à la linéarité conduit à rejeter le plan extrême, optimal du point de vue variance, qui place les x_i aux extrémités de l'intervalle étudié.