

TACTIQUES EN ANALYSE DE VARIANCE ET EN REGRESSION

André Kobilinsky

INRA Route de Saint Cyr

78000 Versailles

Tel:30833352

Résumé :

Après avoir donné en analyse de variance et covariance une définition précise des effets, indépendante des données recueillies, on montre l'impact de la non orthogonalité sur la précision d'estimation de ces effets. Des exemples illustrent la nécessité dans le cas non orthogonal d'étudier des combinaisons linéaires de ces effets, déterminées a posteriori en fonction des données recueillies. Une technique d'obtention de fonctions estimables "simples" est aussi proposée.

Mots-clés: Analyse de variance, covariance, modèle linéaire, fonctions estimables, polynomes orthogonaux

Après des rappels sur les différents types de modèles linéaires et l'inférence statistique dans ces modèles (chap. 1 et 2), on définit les effets qui intéressent le praticien de l'analyse de variance (chap. 3). Lorsque les données observées donnent une non orthogonalité accusée, on ne peut hélas s'arrêter à l'examen de ces seuls effets, qui peuvent être alors mal estimés ou pas du tout estimables. Les chapitres 4 et 5 proposent quelques tactiques simples de rechange pour ces cas non orthogonaux.

1. LES DIFFERENTS TYPES DE MODELE LINEAIRE

Les modèles linéaires décrivent la relation entre une variable quantitative, appelée **variable expliquée** ou **réponse**, et une ou plusieurs variables qualitatives ou quantitatives appelées **variables explicatives** ou **facteurs**. On distingue parmi eux les modèles de **régression** dans lesquels les variables explicatives sont toutes quantitatives, les modèles d'**analyse de variance** où les variables explicatives sont toutes qualitatives et les modèles d'**analyse de covariance** où l'on a un mélange des deux types de variables.

En régression multiple, le modèle s'écrit :

$$y = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k + \epsilon \quad (1.1)$$

On fait l'hypothèse que ϵ est une variable aléatoire d'espérance nulle, de variance σ^2 indépendante des valeurs x_1, \dots, x_k prises par les variables explicatives :

$$\text{var}(\epsilon) = \sigma^2 \quad (1.2)$$

Si l'on note $\theta = (\alpha_0, \alpha_1, \dots, \alpha_k)$ le vecteur colonne des paramètres et $x = (1, x_1, \dots, x_k)$, l'espérance de la réponse y s'écrit :

$$E(y) = \theta x \quad (1.3)$$

Elle dépend linéairement du vecteur θ , et c'est de là que vient l'appellation de modèle linéaire. Un modèle de régression polynomiale tel que :

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \epsilon$$

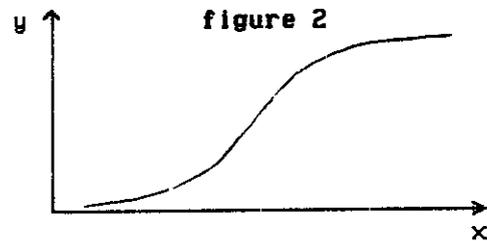
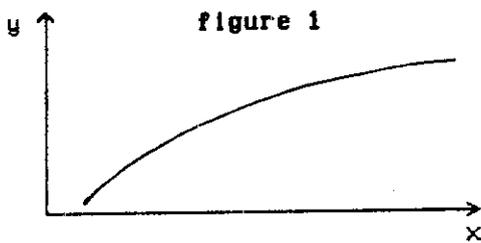
est donc aussi un modèle linéaire, alors même que ce modèle permet de décrire une relation non linéaire entre y et x comme celle de la figure 1. En revanche, le modèle logistique :

$$y = \frac{\alpha_0}{1 + \exp(-\alpha_2(x-\alpha_1))} + \epsilon$$

qui permet de décrire une courbe comme celle de la figure 2 est non linéaire en fonction des paramètres. Un modèle tel que :

$$y = \alpha_0 + \alpha_1 x_1 + (1/\alpha_1) x_2 + \epsilon$$

où la dépendance entre y et x_1, x_2 est linéaire est également un modèle non linéaire, car il dépend de façon non linéaire de α_1 .



Considérons maintenant une expérience de comparaison de différents aliments pour porcs. Il y a une variable explicative qualitative : l'aliment. La variable expliquée y est l'accroissement de poids sur une période allant du sevrage à l'abattage, qui s'écrit pour un aliment "a" donné :

$$y = \mu_a + \epsilon \tag{1.4}$$

μ_a est l'accroissement moyen pour l'aliment a. Comme en régression, on postule que ϵ est d'espérance nulle, de variance σ^2 indépendante de l'aliment a considéré. On peut réécrire (1.4) sous une forme identique à celle du modèle de régression en introduisant les variables indicatrices des k différents aliments :

$$y = \mu_1 x_1 + \dots + \mu_k x_k + \epsilon \tag{1.5}$$

Si l'animal a reçu l'aliment a, les indicatrices x_1, \dots, x_k sont nulles à l'exception de x_a égale à 1.

Dans une telle expérience, on améliore sensiblement la valeur prédictive du modèle en rajoutant une variable explicative telle que le poids p de l'animal au sevrage. Le modèle s'écrit alors :

$$y = \mu_a + \lambda p + \epsilon \quad (1.6)$$

ou encore :

$$y = \mu_1 x_1 + \dots + \mu_k x_k + \lambda p + \epsilon \quad (1.7)$$

La variable explicative quantitative p est généralement qualifiée de **covariable** et le modèle (1.6) qui contient simultanément un facteur qualitatif l'aliment et un facteur quantitatif le poids au sevrage est appelé **modèle d'analyse de covariance**.

Si l'expérience porte sur plusieurs races de porcs, l'accroissement de poids y est a priori fonction de la race et de l'aliment et s'écrit donc (en l'absence de covariable) :

$$y = \mu_{a,r} + \epsilon \quad (1.8)$$

où "a" est l'indice de l'aliment, "r" l'indice de la race. Si les races étudiées ne diffèrent pas trop, l'aliment influe sur la croissance de la même façon pour les différentes races. Le modèle prend alors la forme suivante :

$$y = \alpha_a + \beta_r + \epsilon \quad (1.9)$$

(1.9) est appelé **modèle additif**, par opposition au modèle plus général (1.8) que l'on appelle **modèle interactif**.

2. ESTIMATION DES PARAMETRES

Chacun des modèles précédents dépend de paramètres inconnus (les μ , α , β et λ). Pour estimer ces paramètres, on fait différentes observations de y en faisant varier les valeurs des facteurs. Le modèle donne alors pour chacune des n observations une équation entre les paramètres inconnus. Par exemple en régression multiple, si y_i , x_{1i} , ..., x_{ki} sont les valeurs des différentes variables pour la i ème observation, on a :

$$\begin{aligned} y_1 &= \alpha_0 + \alpha_1 X_{11} + \dots + \alpha_k X_{k1} + \epsilon_1 \\ &\dots \\ y_i &= \alpha_0 + \alpha_1 X_{1i} + \dots + \alpha_k X_{ki} + \epsilon_i \\ &\dots \\ y_n &= \alpha_0 + \alpha_1 X_{1n} + \dots + \alpha_k X_{kn} + \epsilon_n \end{aligned} \tag{2.1}$$

Cet ensemble d'équations peut être écrit sous la forme matricielle :

$$y = X \theta + \epsilon \tag{2.2}$$

où $\theta = (\alpha_0, \alpha_1, \dots, \alpha_k)$ est le vecteur des paramètres.

On s'arrange généralement pour que les observations soient faites indépendamment. Le vecteur ϵ a alors une matrice de covariance diagonale avec le terme σ^2 constant sur la diagonale :

$$\text{cov}(\epsilon) = \sigma^2 I_n \tag{2.3}$$

Dans l'expérience sur les porcs, supposons que l'on expérimente 2 aliments, chacun sur 2 porcs. Si l'aliment 1 est affecté par tirage au sort aux porcs 1 et 4, l'aliment 2 aux porcs 2 et 3, on obtient les équations suivantes :

$$\begin{aligned} y_1 &= \mu_1 + \epsilon_1 \\ y_2 &= \mu_2 + \epsilon_2 \\ y_3 &= \mu_2 + \epsilon_3 \\ y_4 &= \mu_1 + \epsilon_4 \end{aligned} \tag{2.4}$$

qui peuvent se mettre également sous la forme matricielle (2.2), en posant :

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \theta = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

On suppose également l'indépendance entre les réponses, qui entraîne que ϵ a la matrice de covariance donnée en (2.3).

Des modèles tels que (1.6), (1.8) et (1.9) se mettent aussi sous la forme matricielle (2.2), avec la matrice de covariance des erreurs (2.3) lorsqu'il y a indépendance entre les observations.

A partir du système d'équations (2.2), on estime les paramètres et certaines formes linéaires $'a\theta$ de ces paramètres. Avant de voir comment sont choisies ces formes linéaires, nous rappelons les principaux résultats sur l'estimation et les tests.

Dans (2.2), le vecteur y appartient à l'espace vectoriel $F=\mathbb{R}^n$, qui est appelé **espace des observations**. Cet espace est muni du produit scalaire usuel noté $\langle x, y \rangle : \langle x, y \rangle = 'x y$. On note P l'opérateur de projection orthogonale sur l'espace $\text{Im } X$ engendré par les colonnes de la matrice X . Si $('X X)^-$ est une inverse généralisée de $'X X$, on a : $P = X('X X)^-X$. On définit alors les notions suivantes :

a/ Valeurs estimées, résidus

L'estimateur des moindres carrés de $E(y)=X\theta$ est la projection orthogonale P_y de y sur $\text{Im } X$. P_y est aussi appelé le vecteur des **valeurs estimées**, et la différence $y-P_y$ s'appelle le vecteur des **résidus**.

b/ Somme des carrés des résidus

La qualité de l'ajustement peut être mesurée par le carré de la norme du vecteur des résidus : $\langle y-P_y, y-P_y \rangle$. On le désignera en abrégé par **SCR**, ce qui est l'abréviation de **Somme des Carrés des Résidus**.

c/ Carré moyen résiduel

Le rapport $\hat{\theta}^2 = \text{SCR}/\text{dlr}$, où $\text{dlr} = n - \text{rang}(X)$, est un estimateur sans biais de σ^2 . Ce rapport est appelé **Carré Moyen Résiduel** -en abrégé **CMR**- ou encore **variance résiduelle**. Le dénominateur dlr de ce rapport est appelé **nombre de degrés de liberté résiduels**.

d/ Fonction estimable des paramètres

La forme linéaire $'a\theta$ du vecteur θ des paramètres est estimable si on peut l'estimer sans biais à partir d'une forme linéaire $'by$ du vecteur y des observations. Les conditions suivantes sont équivalentes :

- i) $'a\theta$ est estimable
- ii) il existe $b \in \mathbb{R}^n$ tel que $'a\theta = 'bX\theta$ pour tout θ
- iii) $a \in \text{Im } 'X$
- iv) il existe τ tel que $'a\theta = \langle X\tau, X\theta \rangle$ pour tout θ .

L'équivalence de i), ii) et iii) est immédiate, ainsi que l'implication iv) \implies ii). Pour montrer ii) \implies iv), on choisit un vecteur τ tel que $X\tau = Pb$.

Si $'a\theta = \langle X\tau, X\theta \rangle$ pour tout θ , l'estimation sans biais de moindre variance de $'a\theta$ est $\langle X\tau, y \rangle$ qui a pour variance $\sigma^2 \langle X\tau, X\tau \rangle$. Sous l'hypothèse de normalité des erreurs, l'intervalle de confiance de $'a\theta$ au seuil $1-\alpha$ est alors :

$$'a\theta = \langle X\tau, y \rangle \pm t_{\alpha} \hat{\sigma} \sqrt{\langle X\tau, X\tau \rangle}$$

où t_{α} est la valeur qui a la probabilité α d'être dépassée en valeur absolue par un t de student ayant même nombre de degrés de liberté (dlr) que $\hat{\sigma}^2$.

Remarque : il arrive souvent que les réponses soient des moyennes calculées avec des effectifs différents. Dans ce cas, (2.3) est remplacé par :

$$\text{cov}(\epsilon) = \sigma^2 \Pi^{-1} \tag{2.5}$$

où Π est la matrice diagonale des effectifs et σ^2 la variance pour une observation "individuelle". La matrice identité I_n de (2.3) est donc simplement remplacée par une matrice Π^{-1} connue. Dans ce cas, tout ce

qui précède s'applique à condition de remplacer le produit scalaire usuel $\langle x, y \rangle = 'xy$ par le produit scalaire $\langle x, y \rangle = 'x\Pi y$.

e/ Test d'hypothèse

Supposons que l'on veuille tester l'hypothèse H_0 de nullité d'un ensemble de r formes linéaires de θ :

$$H_0 : 'a_1\theta = 0, \dots, 'a_r\theta = 0 \tag{2.6}$$

Notons E_0 le sous espace des vecteurs $X\theta$ tels que θ satisfait H_0 , et P_0 l'opérateur de projection orthogonale sur E_0 . Pour tester l'hypothèse H_0 , on utilise la quantité $SCE = \langle (P - P_0)y, (P - P_0)y \rangle$, appelée **Somme des Carrés des Ecart**s dûe aux r formes linéaires $'a_1\theta, \dots, 'a_r\theta$. Le nombre de degrés de liberté correspondant dl est le rang du projecteur $P - P_0$, qui est égal à la différence entre la dimension de $Im X$ et la dimension du sous-espace E_0 . Le rapport $CM = SCE/dl$ est appelé "**Carré Moyen**", ou encore "**variance**" correspondant à ces r formes. Sous les hypothèses de normalité, le ratio $F = CM/CMR$ suit la loi de Fisher Snedecor, centrée sous l'hypothèse H_0 . Cette dernière hypothèse est donc rejetée si $F > F_\alpha$, où F_α est la valeur qui a la probabilité α d'être dépassée par un F de Fisher Snedecor centré à dl et dlr degrés de liberté.

Pour expliciter la dépendance entre la somme des carrés des écarts SCE et les formes linéaires $'a_1\theta, \dots, 'a_r\theta$, on a besoin de la proposition suivante, qui montre que seules influent sur la statistique de test les fonctions estimables. On note $\{a_1, \dots, a_r\}$ le sous espace engendré par a_1, \dots, a_r , et T le sous espace des formes linéaires associées, c'est à dire des formes $'a\theta$ telles que $a \in \{a_1, \dots, a_r\}$. Parmi ces formes, ne sont estimables que celles pour lesquelles $a \in Im X$. Le sous espace E_0 associé à l'hypothèse nulle vérifie alors :

Proposition . E_0 est l'ensemble des vecteurs $X\theta$ tels que θ annule les formes linéaires estimables de T :

$$E_0 = \{ X\theta / 'a\theta = 0 \text{ pour tout } a \in \{a_1, \dots, a_r\} \cap Im X \}$$

Démonstration : soit $T_e = \{a_1, \dots, a_r\} \cap \text{Im}'X$, T_s un supplémentaire de T_e dans $\{a_1, \dots, a_r\}$ et S un supplémentaire de $\text{Im}'X$ qui contient T_s . $p=k+1$ étant la dimension de \mathfrak{e} , \mathfrak{R}^p est somme directe de $\text{Im}'X$ et S et est donc aussi somme directe de $\text{Ker } X = (\text{Im}'X)^\perp$ et de S^\perp .

Soit alors \mathfrak{e} un vecteur tel que $'a\mathfrak{e} = 0$ pour tout $a \in T_e$. On a : $\mathfrak{e} = \mathfrak{e}_0 + \mathfrak{e}_1$ où $\mathfrak{e}_0 \in \text{Ker } X$, $\mathfrak{e}_1 \in S^\perp$. Par suite $X\mathfrak{e} = X\mathfrak{e}_1$. Or $\mathfrak{e}_1 = \mathfrak{e} - \mathfrak{e}_0$ est orthogonal à T_e puisque \mathfrak{e} l'est par hypothèse et que $\mathfrak{e}_0 \in (\text{Im}'X)^\perp$. D'autre part \mathfrak{e}_1 est orthogonal à S , donc à T_s , et \mathfrak{e}_1 vérifie donc l'hypothèse $H_0 : 'a_1\mathfrak{e}_1 = 0, \dots, 'a_r\mathfrak{e}_1 = 0$. $X\mathfrak{e} = X\mathfrak{e}_1$ appartient donc à E_0 , ce qui démontre la proposition ■

Soient τ_1, \dots, τ_q des vecteurs tels que $'X\tau_1, \dots, 'X\tau_q$ engendrent $\{a_1, \dots, a_r\} \cap \text{Im}'X$. Les formes linéaires estimables de T sont engendrées par $\langle X\tau_1, X\mathfrak{e} \rangle, \dots, \langle X\tau_q, X\mathfrak{e} \rangle$. D'après la proposition, E_0 est le sous espace des vecteurs $X\mathfrak{e}$ qui vérifient :

$$H_0' : \langle X\tau_1, X\mathfrak{e} \rangle = 0, \dots, \langle X\tau_q, X\mathfrak{e} \rangle = 0$$

$P - P_0$ est donc l'opérateur de projection orthogonale sur l'espace $\{X\tau_1, \dots, X\tau_q\}$ engendré par $X\tau_1, \dots, X\tau_q$.

3. DEFINITION DES EFFETS

La théorie résumée dans le paragraphe précédant suppose définies les fonctions des paramètres, ou famille de fonctions, dont on veut tester la nullité ou obtenir des intervalles de confiance. Nous allons voir comment ces fonctions, qu'on appellera également "effets" s'introduisent de façon naturelle à partir du type de modèle considéré.

3.1. Régression multiple

Examinons pour commencer le modèle de régression multiple :

$$y = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k + \epsilon \tag{3.1}$$

Ce modèle peut être utilisé pour faire de la prédiction. La réponse moyenne (espérance de y) pour des valeurs x_1, \dots, x_k des facteurs

$\alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k$ est estimée par $\hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \dots + \hat{\alpha}_k x_k$. Cette prédiction peut être assortie d'un intervalle de confiance qui prend en compte l'imprécision sur l'estimation de $\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_k$. Si l'on désire une prédiction individuelle, on doit aussi prendre en compte la variance de ϵ dans le calcul de l'intervalle de confiance.

On s'intéresse aussi en régression aux paramètres $\alpha_1, \dots, \alpha_k$. Le coefficient α_j de x_j est l'augmentation moyenne de la réponse, lorsqu'on incrémente x_j d'une unité en maintenant à une valeur constante les autres facteurs. C'est ce qu'on appellera l'effet de x_j , ajusté pour les autres facteurs. Si l'hypothèse $\alpha_j=0$ est rejetée par un test, il y a donc dépendance entre la réponse et x_j , toutes choses égales par ailleurs.

Si le domaine de variation des facteurs ne recouvre pas la valeur 0, le coefficient α_0 qui est la réponse moyenne pour $x_1=0, \dots, x_k=0$ n'a pas de sens. On reparamétrise alors (3.1) sous la forme :

$$y = \alpha'_0 + \alpha_1 (x_1 - \bar{x}_1) + \dots + \alpha_k (x_k - \bar{x}_k) + \epsilon \quad (3.2)$$

où $\bar{x}_1, \dots, \bar{x}_k$ sont des valeurs centrales des domaines de variation des facteurs x_1, \dots, x_k et $\alpha'_0 = \alpha_0 + \alpha_1 \bar{x}_1 + \dots + \alpha_k \bar{x}_k$ est la réponse moyenne associée à ces valeurs centrales. Ce dernier paramètre est interprétable contrairement à α_0 .

Si les domaines de variation des facteurs x_1, \dots, x_k ont des largeurs très différentes, il est difficile de comparer les effets $\alpha_1, \dots, \alpha_k$ entre eux. Il est alors judicieux de diviser chaque facteur x_j par une mesure σ_j de l'amplitude de variation de ce facteur. Le modèle (3.2) devient alors :

$$y = \alpha'_0 + \beta_1 (x_1 - \bar{x}_1) / \sigma_1 + \dots + \beta_k (x_k - \bar{x}_k) / \sigma_k + \epsilon$$

Les nouveaux effets $\beta_j = \sigma_j \alpha_j$ s'interprètent de façon analogue aux α_j et ont l'avantage d'être comparables entre eux. On dit qu'ils sont normalisés.

3.2. Régression polynomiale

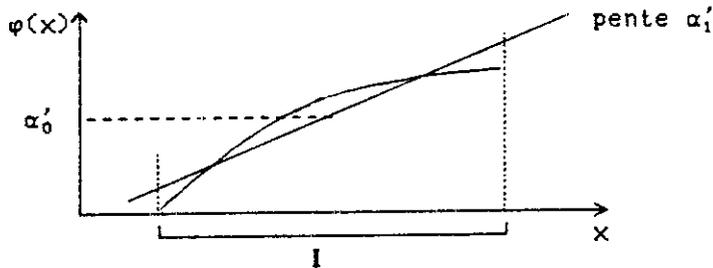
Considérons maintenant le cas de la régression polynomiale :

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \epsilon \quad (3.3)$$

La dérivée de l'espérance est $\alpha_1 + 2\alpha_2 x$. Elle augmente de $2\alpha_2$ lorsque x augmente d'une unité. α_2 s'interprète donc comme une mesure de la courbure. On l'appelle l'effet quadratique. En revanche, α_1 et α_0 , qui dépendent de l'origine choisie pour mesurer la variable explicative x sont généralement ininterprétables. On les remplace par 2 nouveaux paramètres α'_0 et α'_1 définis comme suit :

* α'_0 est la constante qui approxime le mieux la réponse espérée sur le domaine de variation de x considéré. On l'appelle effet moyen.

* α'_1 est la pente de la droite qui approxime le mieux la réponse espérée sur le domaine de variation de x considéré. On l'appelle effet linéaire.



Pour définir ces effets précisément, on introduit un domaine de variation I et une probabilité μ sur ce domaine. La qualité de l'approximation est définie par la distance dans $L^2(\mu)$ à la réponse espérée $\varphi(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2$. Il est alors facile de montrer que l'espérance $\varphi(x)$ de y se réécrit en fonction des nouveaux paramètres α'_0 et α'_1 sous la forme :

$$E(y) = \varphi(x) = \alpha'_0 P_0(x) + \alpha'_1 P_1(x) + \alpha_2 P_2(x) \quad (3.4)$$

où P_0 , P_1 et P_2 sont les polynomes orthogonaux de degré 0, 1, 2 dans $L_2(\mu)$. Si la mesure μ est symétrique autour d'une valeur centrale \bar{x} , ces polynomes s'écrivent :

$$P_0(x) = 1 \quad P_1(x) = x - \bar{x} \quad P_2(x) = (x - \bar{x})^2 - \overline{(x - \bar{x})^2}$$

avec :

$$\bar{x} = \int_I x \, d\mu(x) \quad , \quad \overline{(x - \bar{x})^2} = \int_I (x - \bar{x})^2 \, d\mu(x)$$

Si on note $[\varphi, \psi]$ le produit scalaire de $L_2(\mu)$:

$$[\varphi, \psi] = \int_I \varphi(x) \psi(x) d\mu(x) \quad (3.5)$$

on déduit de (3.4) une expression directe des effets en fonction de φ :

$$\alpha'_0 = [P_0, \varphi]/[P_0, P_0] , \quad \alpha'_1 = [P_1, \varphi]/[P_1, P_1] , \quad \alpha_2 = [P_2, \varphi]/[P_2, P_2]$$

On peut aussi exprimer ces effets comme formes linéaires des paramètres $\alpha_0, \alpha_1, \alpha_2$ soit en utilisant les égalités ci-dessus, soit en développant (3.4). Mais cela est généralement inutile, car α'_0, α'_1 et α_2 sont en pratique estimés directement à partir de la forme (3.4) du modèle.

Normalisation : de même qu'en régression multiple, il est souhaitable pour comparer l'effet linéaire et l'effet quadratique de normer les polynômes orthogonaux associés pour qu'ils aient des variations comparables sur le domaine I considéré. On posera donc :

$$\begin{aligned} R_0 &= P_0/\sqrt{[P_0, P_0]} & R_1 &= P_1/\sqrt{[P_1, P_1]} & R_2 &= P_2/\sqrt{[P_2, P_2]} \\ B_0 &= \alpha'_0 \sqrt{[P_0, P_0]} & B_1 &= \alpha'_1 \sqrt{[P_1, P_1]} & B_2 &= \alpha_2 \sqrt{[P_2, P_2]} \end{aligned}$$

de sorte que :

$$E(y) = \varphi(x) = B_0 R_0(x) + B_1 R_1(x) + B_2 R_2(x)$$

R_0, R_1 et R_2 sont les polynômes orthonormés de degré 0, 1 et 2.

B_0, B_1 et B_2 sont les effets moyen, linéaire et quadratique normalisés. Ils s'expriment en fonction de φ par les relations :

$$B_0 = [R_0, \varphi] \quad B_1 = [R_1, \varphi] \quad B_2 = [R_2, \varphi]$$

Généralisation . Le fait que l'espérance $\varphi(x)$ de y est un polynôme de degré 2 en x n'a pas été utilisé dans la définition des effets moyen et linéaire. On peut donc généraliser la définition de ces effets à n'importe quelle fonction φ dans $L_2(\mu)$ et on peut définir de façon similaire des effets quadratique, cubique, etc... De façon générale, on définit l'effet B_s normalisé de degré s par l'égalité :

$$\beta_s = [R_s, \varphi] \quad (3.6)$$

Nous donnons en annexe l'expression des polynomes orthonormés R_s de degré $s \leq 4$, dans le cas où le domaine I et la probabilité μ sont symétriques autour de 0, cas auquel on peut souvent se ramener par simple changement d'origine sur la variable explicative.

Dans le contexte expérimental, on choisit souvent d'observer la réponse y en des points régulièrement espacés x_1, x_2, \dots, x_r du domaine de variation de x étudié. Dans ce cas là, on choisit généralement pour μ la probabilité discrète donnant une masse égale $1/r$ à chacun de ces points. L'effet β_s devient alors :

$$\beta_s = [\sum_i \varphi(x_i) R_s(x_i)] / r$$

L'intérêt de ce choix est le suivant. Si l'on dispose d'observations de y en chacun des points x_1, \dots, x_r , on peut estimer les espérances $\varphi(x_1), \dots, \varphi(x_r)$ et en déduire l'estimation de β_s . Si l'on note $\hat{\varphi}_i$ l'estimation de $\varphi(x_i)$, on a :

$$\hat{\beta}_s = [\sum_i \hat{\varphi}_i R_s(x_i)] / r$$

Tous les effets sont donc estimables, quelle que soit la forme du modèle de l'espérance φ , qui peut même ne pas être paramétré.

Lorsque les valeurs de x étudiées ne sont pas complètement maîtrisées (données d'enquête), on peut de façon analogue adopter pour μ une probabilité donnant à chaque valeur x un poids égal au nombre d'observations en x . Mais la définition des effets dépend alors de la répartition des points d'observation et il est difficile de comparer des estimations provenant d'échantillons distincts, puisqu'elles n'estiment pas strictement la même chose. Il est donc préférable d'adopter une mesure μ indépendante des données, autrement dit de définir les effets moyen, linéaire, quadratique, ... indépendamment de l'échantillon particulier considéré. Mais il devient alors impératif de spécifier une forme paramétrique pour l'espérance φ si l'on veut que les effets soient estimables.

3.3. Régression polynomiale multiple

On peut adopter une approche similaire à la précédente pour définir des effets en régression polynomiale multiple. Supposons par exemple qu'il y a 2 variables explicatives x_1 et x_2 , que la réponse espérée est $E(y) = \varphi(x_1, x_2)$ et que l'on a choisi des probabilités μ_1 et μ_2 sur les domaines de variation respectifs de x_1 et x_2 . On peut alors définir un effet normalisé $B_{s,t}$ de degré s en x_1 et t en x_2 par une formule analogue à (3.6) :

$$B_{s,t} = [R_s \otimes Q_t, \varphi] = \int \int R_s(x_1) Q_t(x_2) \varphi(x_1, x_2) d\mu_1(x_1) d\mu_2(x_2) \quad (3.7)$$

ou les polynomes R_0, R_1, \dots sont les polynomes orthonormés de degré 0, 1, ... dans $L_2(\mu_1)$ et Q_0, Q_1, \dots de même les polynomes orthonormés de degrés 0, 1, ... dans $L_2(\mu_2)$. Les polynomes produits tensoriels $R_s \otimes Q_t$, définis par :

$$R_s \otimes Q_t (x_1, x_2) = R_s(x_1) Q_t(x_2) \quad (3.8)$$

sont alors les polynomes orthonormés dans $L_2(\mu_1 \otimes \mu_2)$ où $\mu_1 \otimes \mu_2$ est la mesure produit.

Lorsque s et t sont petits, les effets $B_{s,t}$ peuvent être dénommés simplement :

$s=1, t=0$	B_{10} est l'effet lin x_1 (effet linéaire de x_1)
$s=1, t=1$	B_{11} est l'effet lin $x_1 \times$ lin x_2
$s=1, t=2$	B_{12} est l'effet lin $x_1 \times$ quad x_2
	etc ...

Pour interpréter un effet tel que lin $x_1 \times$ lin x_2 , on remarque que le calcul de cet effet peut être décomposé en 2 étapes :

a/ on calcule pour chaque x_2 fixé l'effet $\psi(x_2)$ linéaire de x_1 , c'est à dire la pente de la droite qui approxime le mieux la fonction $x_1 \mapsto \varphi(x_1, x_2)$:

$$\psi(x_2) = \int R_1(x_1) \varphi(x_1, x_2) d\mu_1(x_1)$$

b/ on calcule ensuite l'effet linéaire pour la réponse $\psi(x_2)$, c'est à dire la pente de la droite qui approxime le mieux $\psi(x_2)$

$$B_{11} = \int R_2(x_2) \psi(x_2) d\mu(x_2)$$

Cet effet est ce qu'on appelle une mesure d'interaction entre les facteurs x_1 et x_2 . S'il est important, cela indique que la pente moyenne de la relation liant la réponse à x_1 a une variation monotone et régulière en fonction de x_2 .

L'utilisation d'une mesure produit $\mu_1 \otimes \mu_2$ est ici essentielle si l'on veut une interprétation simple des effets considérés. Dans le cas de données d'enquête, la mesure donnant à chaque point (x_1, x_2) un poids proportionnel au nombre d'observations en ce point n'est pas une mesure produit et n'est pas adéquate pour définir des effets.

3.4. Analyse de variance à 1 facteur

Dans l'expérience sur les porcs, on est amené pour comparer les aliments entre eux à étudier des formes linéaires telles que :

$$\begin{aligned} \mu_2 - \mu_1 & \quad (\text{comparaison entre les aliments 2 et 1}) \\ \mu_3 - (\mu_1 + \mu_2)/2 & \quad (\text{comparaison entre l'aliment 3 et la moyenne des aliments 1 et 2}) \end{aligned}$$

Plus généralement on étudie des formes linéaires $\sum_a c_a \mu_a$ des espérances μ_a dont la somme des coefficients est nulle : $\sum_a c_a = 0$. De telles formes linéaires sont appelées des contrastes de l'effet aliment. La famille de tous ces contrastes, qui est engendrée par les différences $\mu_a - \mu_{a'}$ entre 2 aliments a et a' distincts constitue "l'effet principal du facteur aliment".

On a souvent l'habitude de réécrire le modèle (1.4) sous la forme :

$$y = \mu + \alpha_a + \epsilon \tag{3.9}$$

où : $\mu = \mu_{..} = (\sum_a \mu_a) / k$ (k : nombre d'aliments comparés)

$$\alpha_a = \mu_a - \mu$$

α_a est appelé effet de l'aliment a. C'est un contraste de l'effet principal aliment et les différents α_a engendrent cet effet.

3.5. Analyse de variance à 2 facteurs

Si l'expérience porte sur plusieurs races, on se pose la question de savoir si les différences entre aliments dépendent ou pas de la race. Cette question peut être formalisée par l'étude d'un contraste tel que :

$$C = (\mu_{a' r'} - \mu_{a r'}) - (\mu_{a' r} - \mu_{a r}) \quad (3.10)$$

C est nul si la différence entre les aliments a' et a est la même pour les races r et r'. On appelle contrastes d'interaction entre les facteurs aliment et race les formes linéaires des μ_{ar} générées par l'ensemble des contrastes C associés aux différents quadruplets (a,a',r,r'). L'ensemble de ces contrastes est ce qu'on appelle l'interaction Aliment x Race (on parlera aussi d'effet d'interaction).

Si les contrastes d'interaction ne sont pas trop grands et que l'effet aliment est donc du même ordre pour les différentes races, on calcule des différences moyennes telles que :

$$A = \mu_{a..} - \mu_{a'..} \quad (3.11)$$

L'ensemble de ces différences engendre l'effet principal du facteur aliment.

L'interaction définie ci-dessus donne un rôle symétrique aux 2 facteurs. Cela apparaît clairement si l'on réécrit C sous la forme suivante :

$$C = (\mu_{a' r'} - \mu_{a r'}) - (\mu_{a r'} - \mu_{a r})$$

Si les contrastes d'interaction sont petits, les différences entre races seront donc également du même ordre quel que soit l'aliment utilisé et on pourra les résumer par des différences moyennes comme :

$$R = \mu_{..r'} - \mu_{..r} \quad (3.12)$$

L'ensemble de ces différences engendre l'effet principal race.

Les moyennes $\mu_{a..}$ et $\mu_{.r}$ sont généralement non pondérées :

$$\mu_{a..} = (\mu_{a1} + \dots + \mu_{ae}) / \ell \quad (\ell : \text{nb. de races})$$

$$\mu_{.r} = (\mu_{1r} + \dots + \mu_{kr}) / k \quad (k : \text{nb. d'aliments})$$

Mais on peut aussi choisir de les définir comme des moyennes pondérées :

$$\mu_{a..} = p_1 \mu_{a1} + \dots + p_e \mu_{ae} \quad , \quad p_1 + \dots + p_e = 1 \quad (3.13)$$

$$\mu_{.r} = q_1 \mu_{1r} + \dots + q_k \mu_{kr} \quad , \quad q_1 + \dots + q_k = 1 \quad (3.14)$$

Par exemple, les poids p_1, \dots, p_e sont choisis proportionnels aux nombres de porcs de chaque race dans la zone à laquelle s'applique l'étude.

De par la liberté de choix des poids et de même qu'en régression polynomiale, il entre donc une certaine part d'arbitraire dans la définition des effets principaux. Cependant, pour qu'une différence entre moyennes $\mu_{a..} - \mu_{a'..}$ soit aussi une moyenne des différences $\mu_{a.r} - \mu_{a'.r}$, il est impératif que les poids p_1, \dots, p_e affectés aux races pour le calcul des moyennes $\mu_{a..}$ soient les mêmes pour tous les aliments.

De même qu'en analyse de variance à un facteur, on a l'habitude de réécrire le modèle (1.8) sous la forme :

$$y = \mu + \alpha_a + \beta_r + \gamma_{ar} + \epsilon \quad (3.15)$$

$$\text{où : } \mu = \mu_{...} = \sum_{a,r} q_a p_r \mu_{a.r} = \sum_a q_a \mu_{a..}$$

$$\alpha_a = \mu_{a..} - \mu_{...}$$

$$\beta_r = \mu_{.r} - \mu_{...}$$

$$\gamma_{ar} = \mu_{a.r} - \mu_{a..} - \mu_{.r} + \mu_{...}$$

Notons que ces paramètres vérifient les contraintes :

$$\sum_a \gamma_{ar} = 0 \quad \sum_r \gamma_{ar} = 0 \quad \sum_a \alpha_a = 0 \quad \sum_r \beta_r = 0 \quad (3.16)$$

Les γ_{ar} engendrent l'interaction aliment x race. Les α_a et β_r engendrent respectivement l'effet principal aliment et l'effet principal race.

Définition d'effets par programme

Une fois précisé le modèle qui peut s'écrire symboliquement dans l'exemple ci-dessus :

$$Y = A..R + EPS$$

les contrastes intéressant l'utilisateur s'introduisent comme combinaison linéaire de moyennes. Pour définir une moyenne, il suffit de préciser les facteurs dont le niveau est fixé, comme dans les exemples suivant (cf aussi MODULAD, notice du module MODLI).

$$\begin{array}{ll} M(A=1,R=1) & \text{pour } \mu_{11} \\ M(A=1) & \text{pour } \mu_{1.} = \sum_r p_r \mu_{1r} \\ M(R=1) & \text{pour } \mu_{.1} = \sum_a q_a \mu_{a1} \\ M & \text{pour } \mu_{..} = \sum_{a,r} q_a p_r \mu_{ar} \end{array}$$

On peut alors facilement définir des contrastes tels que les différences $\mu_{a.} - \mu_{.}$, ou les effets α_a et γ_{ar} :

$$\begin{array}{ll} M(A=2) - M(A=1) & \text{pour } \mu_{2.} - \mu_{1.} \\ M(A=1) - M & \text{pour } \alpha_1 = \mu_{1.} - \mu_{..} \\ M(A=1,R=1) - M(A=1) - M(R=1) + M & \text{pour } \gamma_{11} = \mu_{11} - \mu_{1.} - \mu_{.1} + \mu_{..} \end{array}$$

Si l'on veut définir une famille de contrastes, on remplace les indices 1 et 2 figurant ci-dessus par des variables :

$$M(A=A2) - M(A=A1) \text{ pour définir l'ensemble des différences } \mu_{a.} - \mu_{.}$$

Dans certains cas, on peut sous-entendre certains indices. Ainsi :

$$\begin{array}{ll} M(A) - M & \text{définit les effets } \alpha_a = \mu_{a.} - \mu_{..}, \text{ de même que} \\ M(A=A) - M & \text{(l'indice est ici noté comme le facteur)} \end{array}$$

3.6. Définition des effets dans un cadre plus général

(* ce paragraphe peut être omis en première lecture)

Supposons maintenant qu'il y a f facteurs, quantitatifs ou qualitatifs et soient I_1, \dots, I_f les ensembles de niveaux correspondants. I_j est un intervalle de R si le facteur j est

quantitatif, un ensemble fini de niveaux si le facteur j est qualitatif. On note $\varphi(x_1, \dots, x_p)$ l'espérance de la réponse y pour les niveaux x_1, \dots, x_p des facteurs. Le modèle s'écrit donc :

$$y = \varphi(x_1, \dots, x_p) + \epsilon \quad E(\epsilon) = 0 \quad (3.17)$$

φ est supposée définie pour chaque combinaison de niveaux $(x_1, \dots, x_p) \in I_1 \times \dots \times I_p$ des facteurs. D'un point de vue pratique, cela indique qu'il n'y a pas de hiérarchie entre les facteurs. Nous verrons ultérieurement comment adapter la présente démarche lorsqu'il y a des hiérarchies.

Pour définir des effets, on choisit une probabilité μ_j sur chaque intervalle I_j . Si le facteur j est quantitatif, on note $R_{j0}, R_{j1}, R_{j2}, \dots$ les polynômes orthogonaux de degré $0, 1, 2, \dots$ dans $L_2(\mu_j)$ et $E_{j0}, E_{j1}, E_{j2}, \dots$ les sous-espaces de dimension 1 qu'ils engendrent. Si le facteur j est qualitatif, on note E_{j0} le sous-espace engendré par le vecteur constant de $L_2(\mu_j)$ et E_{j1} l'orthogonal de E_{j0} dans $L_2(\mu_j)$. On peut alors associer un effet à chaque p -uplet d'indices (i_1, \dots, i_p) tel que i_j prend les valeurs $0, 1, 2, \dots$ si le facteur j est quantitatif, les valeurs 0 et 1 seulement si le facteur j est qualitatif. Cet effet est par définition le sous-espace $\mathfrak{L}(i_1, \dots, i_p)$ de toutes les formes linéaires de φ qui s'écrivent :

$$\begin{aligned} B(\varphi) &= [R_1 \otimes \dots \otimes R_p, \varphi] \quad (3.18) \\ &= \int \dots \int R_1(x_1) \dots R_p(x_p) \varphi(x_1, \dots, x_p) d\mu_1(x_1) \dots d\mu_p(x_p) \end{aligned}$$

où $R_1 \in E_{1, i_1}, \dots, R_p \in E_{p, i_p}$. Le produit tensoriel $R_1 \otimes \dots \otimes R_p$ est défini comme en régression multiple par :

$$R_1 \otimes \dots \otimes R_p(x_1, \dots, x_p) = R_1(x_1) \dots R_p(x_p) \quad (3.19)$$

C'est un vecteur de $L_2(\mu_1 \otimes \dots \otimes \mu_p)$ et la notation $[,]$ utilisée dans (3.18) fait référence au produit scalaire dans cet espace. On suppose évidemment, pour donner un sens à (3.18), que φ appartient à $L_2(\mu_1 \otimes \dots \otimes \mu_p)$. Par ailleurs, on identifie dans ce qui suit chaque

forme linéaire $B(\varphi)$ au vecteur $R_1 \otimes \dots \otimes R_p$ qui la définit. L'espace $\mathfrak{L}(i_1, \dots, i_p)$ peut être ainsi considéré comme un sous-espace de $L_2(\mu_1 \otimes \dots \otimes \mu_p)$. De par sa définition, on a :

$$\mathfrak{L}(i_1, \dots, i_p) = E_{1, i_1} \otimes \dots \otimes E_{p, i_p} \quad (3.20)$$

L'interprétation de l'effet $B(\varphi)$ est analogue à celle des exemples donnés dans les paragraphes précédents. Si $i_1 = i_2 = \dots = i_p = 0$, $R_1 \dots R_p$ sont des constantes et $B(\varphi)$ est proportionnelle à la moyenne générale $\varphi(\cdot, \dots, \cdot)$:

$$B(\varphi) \propto \varphi(\cdot, \dots, \cdot) = \int \dots \int \varphi(x_1, \dots, x_p) d\mu_1(x_1) \dots d\mu_p(x_p)$$

Si un seul des indices i_1, \dots, i_p est non nul, par exemple i_1 , on a :

$$B(\varphi) \propto \int R_1(x_1) \varphi(x_1, \cdot, \dots, \cdot) d\mu_1(x_1)$$

$$\text{où : } \varphi(x_1, \cdot, \dots, \cdot) = \int_{x_2} \dots \int_{x_p} \varphi(x_1, \dots, x_p) d\mu_2(x_2) \dots d\mu_p(x_p)$$

$B(\varphi)$ est alors un contraste entre les moyennes marginales $\varphi(x_1, \cdot, \dots, \cdot)$ du premier facteur. L'ensemble $\mathfrak{L}(1, 0, \dots, 0)$ de tous ces contrastes constitue l'effet principal du facteur 1 dans le cas qualitatif. Dans le cas quantitatif, il s'agit de l'effet de degré i_1 de ce facteur.

Si deux des indices sont non nuls, disons i_1 et i_2 , on a :

$$B(\varphi) \propto \int R_1(x_1) \int R_2(x_2) \varphi(x_1, x_2, \cdot, \dots, \cdot) d\mu_1(x_1) d\mu_2(x_2)$$

$$\text{où : } \varphi(x_1, x_2, \cdot, \dots, \cdot) = \int_{x_3} \dots \int_{x_p} \varphi(x_1, \dots, x_p) d\mu_3(x_3) \dots d\mu_p(x_p)$$

Ainsi le calcul peut se faire en 2 temps. Pour chaque valeur fixée de x_1 , on calcule un contraste $\delta(x_1, \varphi)$ entre les moyennes marginales $\varphi(x_1, x_2, \cdot, \dots, \cdot)$:

$$\delta(x_1, \varphi) = \int R_2(x_2) \varphi(x_1, x_2, \cdot, \dots, \cdot) d\mu_2(x_2)$$

Puis on forme un nouveau contraste entre ces $\delta(x_1, \varphi)$:

$$B(\varphi) \propto \int R_1(x_1) \delta(x_1, \varphi) d\mu_1(x_1)$$

On dit que le contraste ainsi formé appartient à l'interaction des facteurs 1 et 2. Si le facteur 1 est quantitatif, on précise que $\beta(\varphi)$ est de degré i_1 en x_1 , et de même pour le facteur 2.

Facteurs hiérarchisés

On peut souvent se ramener à la situation précédente en introduisant des pseudo-facteurs. Ainsi dans l'expérience sur les porcs, si on utilise plusieurs animaux par portée, on introduira en sus des facteurs aliment et race, le facteur "portée" qui est hiérarchisé par le facteur race. Si le nombre de portées utilisé par race est constant on peut alors décrire chaque situation possible par un triplet $(x_1, x_2, x_3) = (a, r, p)$ où a est l'aliment, r la race et p le numéro de portée dans la race. Ce triplet appartient à $I_1 \times I_2 \times I_3$ où I_1 est l'ensemble des aliments comparés, I_2 l'ensemble des races et I_3 l'ensemble des indices de répétition, de cardinal égal au nombre de portées par race. La donnée de ce numéro de portée définit ce qu'on appelle un pseudo-facteur. Isolé, ce pseudo-facteur n'a pas de sens puisque 2 portées de 2 races distinctes qui porte le même numéro n'ont pas de relation entre elles. Il faut donc toujours associer ce pseudo-facteur au facteur race, ce qui conduit à regrouper dans la définition des effets le "pseudo" effet principal portée $\mathfrak{z}(0,0,1)$ à la "pseudo" interaction portée \times race $\mathfrak{z}(0,1,1)$. L'ensemble des contrastes générés par ces 2 effets constitue l'effet "portée intra race".

De façon plus générale, supposons pour simplifier que les facteurs sont tous qualitatifs et que I_1, \dots, I_p sont les ensembles de niveaux des facteurs ou pseudo-facteurs, qui sont munis de probabilités μ_1, \dots, μ_p . La relation de hiérarchie définit un préordre partiel sur l'ensemble $\mathcal{F} = \{1, \dots, p\}$ des numéros de facteurs : le facteur j est inférieur au facteur i , et on note $j < i$, si i est hiérarchisé par j (dans l'exemple précédent, on a : race $<$ portée). On associe alors un effet à chaque sous-ensemble \mathcal{A} de \mathcal{F} qui vérifie :

$$i \in \mathcal{A} \text{ et } j < i \implies j \notin \mathcal{A} \quad (3.21)$$

Cet effet est le sous-espace $\mathfrak{z}(\mathcal{A})$ des formes linéaires $\beta(\varphi)$ définies par (3.18) à partir de vecteurs R_1, \dots, R_p vérifiant :

- a/ $R_j \in E_{j1}$ si $j \in \mathcal{A}$
- b/ R_j est un vecteur quelconque de $\mathbb{R}^{|\mathcal{I}_j|}$ si $j < i$ et $i \in \mathcal{A}$
- c/ $R_j \in E_{j0}$ dans tous les autres cas, c'est à dire s'il n'existe pas de $i \in \mathcal{A}$ tel que $j < i$.

L'interprétation de ces effets est simple. Reprenons l'exemple des porcs avec $\mathcal{F}=\{1,2,3\}$ et l'unique relation $2 < 3$. L'ensemble $\mathcal{A}=\{3\}$ vérifie (3.21) et définit donc un effet formé par les contrastes $\mathcal{B}(\varphi)$ de la forme :

$$\mathcal{B}(\varphi) \propto \int R_2(x_2) \int R_3(x_3) \varphi(\cdot, x_2, x_3) d\mu_2(x_2) d\mu_3(x_3)$$

où R_3 est orthogonal dans $L_2(\mu_3)$ au vecteur constant 1_k et R_2 est quelconque. En particulier, si R_2 est égal à 1 pour l'une des races $r=x_2$, à 0 pour les autres, on voit que $\mathcal{B}(\varphi)$ est un contraste entre les moyennes $\varphi(\cdot, x_2, x_3)$ associées aux différentes portées x_3 de race x_2 , et ces contrastes engendrent $\mathcal{L}(\mathcal{A})$ qui est l'effet portée intra-race.

Si certains des facteurs sont quantitatifs, on suppose qu'il n'y a de relation de hiérarchie ni entre eux, ni entre ces facteurs d'une part et les facteurs qualitatifs d'autre part. On peut alors associer un effet à chaque f uplet (i_1, \dots, i_p) pour lequel l'ensemble \mathcal{A} des j tels que le facteur j soit qualitatif et $i_j=1$ vérifie (3.21). Cet effet, noté comme précédemment $\mathcal{L}(i_1, \dots, i_p)$, est l'ensemble des produits $R_1 \otimes \dots \otimes R_p$ tels que :

$$R_j \in E_{j1_j} \quad \text{si le facteur } j \text{ est quantitatif}$$

R_j vérifie les conditions a, b, c données ci-dessus si le facteur j est qualitatif.

Posons :

$F_j = E_{j1_j}$	pour un facteur j quantitatif
$F_j = E_{j1}$	si $j \in \mathcal{A}$
$F_j = \mathbb{R}^{ \mathcal{I}_j }$	si $j < i$ pour $i \in \mathcal{A}$
$F_j = E_{j0}$	si $i_j=0$ et s'il n'existe pas de $i \in \mathcal{A}$ tel que $j < i$

On a alors :

$$\mathfrak{L}(i_1, \dots, i_p) = F_1 \otimes \dots \otimes F_p \quad (3.22)$$

Changement de paramètres

Les effets précédemment introduits sont orthogonaux dans $L_2(\mu_1 \otimes \dots \otimes \mu_p)$. Cela résulte de l'identité :

$$[R_1 \otimes \dots \otimes R_p, S_1 \otimes \dots \otimes S_p] = [R_1, S_1][R_2, S_2] \dots [R_p, S_p] \quad (3.23)$$

qui entraîne l'orthogonalité des 2 vecteurs $R_1 \otimes \dots \otimes R_p$ et $S_1 \otimes \dots \otimes S_p$ dès que $R_j \perp S_j$ pour l'un des indices j . Utilisant cette propriété, on peut aussi donner une base orthogonale simple de l'espace $\mathfrak{L}(i_1, \dots, i_p)$ défini en (3.22). On choisit pour cela une base orthogonale \mathfrak{R}_j dans chaque sous-espace F_j , et on forme les produits tensoriels $R_1 \otimes \dots \otimes R_p$ pour $R_1 \in \mathfrak{R}_1, \dots, R_p \in \mathfrak{R}_p$.

En pratique, on fait souvent l'hypothèse que φ appartient à un sous-espace du type $E = \Sigma \mathfrak{L}(i_1, \dots, i_p)$ de $L_2(\mu_1 \otimes \dots \otimes \mu_p)$, la sommation portant sur une famille \mathfrak{L} d'effets. On peut alors former, à partir de bases orthonormées des sous-espaces orthogonaux $\mathfrak{L}(i_1, \dots, i_p)$, une base orthonormée z_1, \dots, z_p de E tel que chaque z_j appartient à l'un des espaces $\mathfrak{L}(i_1, \dots, i_p)$. Décomposant φ sur cette base, on obtient :

$$\varphi = \Sigma_j [z_j, \varphi] z_j \quad (3.24)$$

(3.24) donne une reparamétrisation en fonction de paramètres $\beta_j = [z_j, \varphi]$ appartenant aux différents effets, donc facilement interprétables.

Néanmoins, lorsqu'il y a des facteurs qualitatifs, on utilise généralement une autre reparamétrisation, analogue à celle donnée en (3.15) dans le cas particulier de l'analyse de variance à 2 facteurs. Les paramètres de cette reparamétrisation sont redondants, mais ont l'avantage de respecter la symétrie entre les niveaux de chaque facteur qualitatif. Pour l'obtenir, on commence par décomposer φ sur les sous-espaces associés aux différents effets de \mathfrak{L} . On note donc P_e l'opérateur de projection orthogonale sur le sous espace $\mathfrak{L}(i_1, \dots, i_p)$ associé à un effet e de \mathfrak{L} , de sorte que :

$$\varphi = \sum_e P_e \varphi \tag{3.25}$$

On décompose ensuite $P_e \varphi$ sur une base orthonormée formée à partir, d'une part des polynômes orthonormés pour les facteurs quantitatifs, d'autre part des bases canoniques de $L_2(\mu_j)$ pour les facteurs qualitatifs. Plus précisément, si le facteur j est qualitatif, on note \mathfrak{X}_j la base orthonormée obtenue en normant les vecteurs canoniques $e_1=(1,0,\dots,0)$, $e_2=(0,1,0,\dots)$... de $L_2(\mu_j)$. Si le facteur j est quantitatif, on note \mathfrak{X}_j la base formée du seul polynôme orthogonal normé de degré i_j . On a alors :

$$P_e \varphi = \sum [R_1 \otimes \dots \otimes R_p, P_e \varphi] R_1 \otimes \dots \otimes R_p \tag{3.26}$$

la sommation étant étendue à l'ensemble $\mathfrak{X}_1 \otimes \dots \otimes \mathfrak{X}_p$ de tous les produits tensoriels $R_1 \otimes \dots \otimes R_p$ tels que $R_1 \in \mathfrak{X}_1, \dots, R_p \in \mathfrak{X}_p$.

Pour expliciter les "nouveaux paramètres", on remarque que : $[R_1 \otimes \dots \otimes R_p, P_e \varphi] = [P_e(R_1 \otimes \dots \otimes R_p), \varphi]$. D'autre part, si $\mathfrak{I}(i_1, \dots, i_p) = F_1 \otimes \dots \otimes F_p$ et si A_1, \dots, A_p sont les opérateurs de projection orthogonale sur F_1, \dots, F_p respectivement, il est facile de montrer en utilisant (3.23) que P_e est le produit tensoriel $P_e = A_1 \otimes \dots \otimes A_p$, défini par :

$$P_e(S_1 \otimes \dots \otimes S_p) = A_1 \otimes \dots \otimes A_p (S_1 \otimes \dots \otimes S_p) = A_1 S_1 \otimes \dots \otimes A_p S_p \tag{3.27}$$

On a donc : $P_e(R_1 \otimes \dots \otimes R_p) = A_1 R_1 \otimes \dots \otimes A_p R_p$. Si le facteur j est quantitatif, $A_j R_j = R_j$. Si il est qualitatif, l'opérateur A_j est soit l'identité I , soit l'opérateur P_0 de projection sur le vecteur constant $\mathbf{1}$, soit l'opérateur $I - P_0$ de projection sur l'orthogonal du vecteur $\mathbf{1}$. $A_j S_j$ est donc égal à S_j , à $P_0 S_j$ ou à $S_j - P_0 S_j$. On pourra alors remplacer chaque terme $A_j S_j$ par le vecteur donné ci-dessus et développer en utilisant la propriété de multilinéarité (évidente d'après la définition) du produit tensoriel.

Si l'on examine une combinaison de niveaux (x_1, \dots, x_p) particulière des facteurs, on obtient l'expression suivante :

$$\varphi(x_1, \dots, x_p) = \sum \sum [R_1 \otimes \dots \otimes R_p, P_e \varphi] R_1 \otimes \dots \otimes R_p(x_1, \dots, x_p) \quad (3.28)$$

On peut noter dans le produit $R_1 \otimes \dots \otimes R_p(x_1 \otimes \dots \otimes x_p) = R_1(x_1) \dots R_p(x_p)$, que les termes associés aux facteurs j qualitatifs sont nuls à l'exception de ceux pour lesquels R_j est précisément le vecteur normé déduit du x_j ème vecteur canonique de $L_2(\mu_j)$. Cela permet en fin de compte d'obtenir une expression simple analogue à (3.15).

Pour illustrer, reprenons l'exemple des porcs avec les facteurs aliment (1) à k niveaux, race (2) à ℓ niveaux et poids au sevrage (3). Pour chacun des 2 facteurs qualitatifs, on choisit de donner le même poids à tous ses niveaux : $\mu_1(a)=1/k$, $\mu_2(r)=1/\ell$. On note e_a , e_r les vecteurs canoniques de R^k et R^ℓ associés respectivement à l'aliment a et à la race r , R_1 et R_2 les vecteurs normés correspondants, R_3 le polynome orthogonal de degré 1 du poids :

$$R_1 = \sqrt{k} e_a \quad R_2 = \sqrt{\ell} e_r \quad R_3(p) \propto p - \bar{p}$$

Enfin on note P_1 , P_2 les opérateurs de projection orthogonale sur les espaces engendrés par $\mathbf{1}_k$ et $\mathbf{1}_\ell$ respectivement :

$$P_1 = \mathbf{1}_k \mathbf{1}_k' / k \quad P_2 = \mathbf{1}_\ell \mathbf{1}_\ell' / \ell$$

Si on considère l'effet e d'interaction "aliment x effet linéaire du poids", on est amené à calculer :

$$[R_1 \otimes R_2 \otimes R_3, P_e \varphi] = [A_1 R_1 \otimes A_2 R_2 \otimes A_3 R_3, \varphi] = [(I - P_1) R_1 \otimes P_2 R_2 \otimes R_3, \varphi] =$$

$$[R_1 \otimes P_2 R_2 \otimes R_3, \varphi] - [P_1 R_1 \otimes P_2 R_2 \otimes R_3, \varphi]$$

$$\text{On a : } P_1 R_1 = (1/\sqrt{k}) \mathbf{1}_k \quad P_2 R_2 = (1/\sqrt{\ell}) \mathbf{1}_\ell$$

En utilisant ces expressions et après intégration partielle, on trouve :

$$[R_1 \otimes P_2 R_2 \otimes R_3, \varphi] = \sqrt{1/k\ell} \int R_3(x_3) \varphi(a, \cdot, x_3) d\mu_3(x_3)$$

$$[P_1 R_1 \otimes P_2 R_2 \otimes R_3, \varphi] = \sqrt{1/k\ell} \int R_3(x_3) \varphi(\cdot, \cdot, x_3) d\mu_3(x_3)$$

$$R_1 \otimes R_2 \otimes R_3(x_1 \otimes x_2 \otimes p) = 0 \quad \text{si } x_1 \neq a \text{ ou } x_2 \neq r$$

$$= \sqrt{k\ell} R_3(p) \quad \text{si } x_1 = a \text{ et } x_2 = r$$

En fin de compte (3.26) devient :

$$(P_e \varphi)(a, r, p) = \delta_a R_3(p)$$

$$\text{avec : } \delta_a = \int R_3(p) (\varphi(a, \cdot, p) - \varphi(\cdot, \cdot, p)) d\mu_3(p)$$

$$\text{On peut noter que } \int \mu_1(a) \delta_a = \int \delta_a d\mu_1(a) = 0$$

δ_a s'interprète comme la différence entre l'effet linéaire du poids pour le régime alimentaire a et l'effet linéaire moyen du poids.

4. Non orthogonalité

La définition des effets donnée précédemment est indépendante de l'expérience ou de l'enquête effectivement réalisée. Aussi il se peut dans certain cas de figure qu'on ne soit pas en mesure d'obtenir d'estimations fiables des effets ainsi définis, alors qu'on peut estimer assez précisément certaines combinaisons linéaires simples de ces effets.

Considérons par exemple le cas d'une régression multiple sur 2 variables explicatives x_1, x_2 . On suppose que dans l'échantillon analysé, il y a un coefficient de corrélation ρ entre x_1 et x_2 . On note z_1, z_2 les variables obtenues par centrage et réduction de x_1 et x_2 . Le modèle associé aux différentes observations s'écrit alors :

$$y = X \tau + \varepsilon$$

$$\text{où : } \tau = \begin{bmatrix} \mu \\ \rho \mu \\ \rho \mu \end{bmatrix} \quad X = \begin{bmatrix} 1 & z_{11} & z_{21} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & z_{1n} & z_{2n} \end{bmatrix} = [1 \ z_1 \ z_2]$$

La matrice tXX des équations normales et son inverse sont donc :

$${}^tXX = n \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix} \quad ({}^tXX)^{-1} = \frac{1}{n} \frac{1}{1-\rho^2} \begin{bmatrix} 1-\rho^2 & 0 & 0 \\ 0 & 1 & -\rho \\ 0 & -\rho & 1 \end{bmatrix}$$

L'estimation des moindres carrés $\hat{\tau} = (\hat{\mu}, \hat{\rho}_1, \hat{\rho}_2)$ a $\sigma^2 ({}^tXX)^{-1}$ pour matrice de covariance. Les variances des estimations $\hat{\rho}_1, \hat{\rho}_2$ des effets linéaires normalisés sont donc égales à $\sigma^2 / (n(1-\rho^2))$. Au dénominateur de cette variance figure la quantité $1-\rho^2$, mesure de la variance de z_1

à z_2 fixé. Si cette variance est trop faible, on ne peut pas estimer correctement l'effet α_1 de z_1 pour z_2 fixé, et de façon symétrique on n'est également dans l'incapacité d'estimer correctement α_2 .

L'impossibilité d'estimer précisément chacun des paramètres α_1 et α_2 , à cause d'un coefficient de corrélation ρ élevé, ne signifie pas pour autant que les données n'apporte pas d'information intéressante sur α_1 et α_2 . En fait, la somme $\alpha_1 + \alpha_2$ est estimée avec la variance :

$$\text{var}(\hat{\alpha}_1 + \hat{\alpha}_2) = \frac{2 \sigma^2}{n (1 + \rho)}$$

qui diminue jusqu'à σ^2/n lorsque ρ croit vers 1. Si l'un des facteurs x_1 ou x_2 a un effet appréciable et que ρ est proche de 1, on trouve que la somme $\alpha_1 + \alpha_2$ est significativement différente de 0, alors qu'aucun des effets α_1 , α_2 ne l'est.

Lorsque $\rho=0$, la matrice 'XX est diagonale et les estimations $\hat{\mu}$, $\hat{\alpha}_1$, $\hat{\alpha}_2$ sont non corrélées. On dit alors qu'il y a orthogonalité des facteurs étudiés. Dans ce cas, l'examen direct des paramètres d'intérêt μ , α_1 , α_2 permet d'extraire toute l'information que les données apportent sur ces paramètres. Lorsqu'au contraire il y a un écart notable à l'orthogonalité ($|\rho|$ proche de 1), on voit qu'il est indispensable d'examiner aussi certaines combinaisons linéaires des paramètres d'intérêt telle que $\alpha_1 + \alpha_2$. Une tactique usuelle pour obtenir de telles combinaisons linéaires consiste à estimer dans le cadre d'un sous-modèle, puis à regarder l'espérance des estimateurs dans le modèle complet. Ainsi, dans l'exemple ci-dessus, après avoir constaté que α_1 et α_2 ne sont pas significativement différents de 0 à cause d'une variance d'estimation importante liée à la non orthogonalité, on estime α_1 dans le cadre du sous-modèle défini par $\alpha_2=0$. Le nouvel estimateur $\tilde{\alpha}_1 = ({}^t z_1 y)/n$, dont on montre aisément qu'il est non corrélé à l'estimateur $\hat{\alpha}_2$ du modèle complet, a pour variance σ^2/n et pour espérance :

$$E(\tilde{\alpha}_1) = \alpha_1 ({}^t z_1 z_1 / n) + \alpha_2 ({}^t z_1 z_2 / n) = \alpha_1 + \rho \alpha_2$$

On peut faire une démarche exactement analogue dans le cas de l'analyse de variance à 2 facteurs. Reprenons l'exemple des porcs avec les 2 facteurs aliment, race (modèle (1.8)). On définit les effets principaux correspondants à partir des moyennes pondérées par les poids q_a et p_r . L'espérance μ_{ar} est estimée par la moyenne $\hat{\mu}_{ar} = \bar{y}_{ar}$ des réponses y_{ar} de porcs de race r élevés avec a . Les estimations des moyennes sont alors :

$$\hat{\mu}_{a..} = \sum_r p_r \bar{y}_{ar} \quad \hat{\mu}_{.r.} = \sum_a q_a \bar{y}_{ar}$$

et : $\text{cov}(\hat{\mu}_{a..}, \hat{\mu}_{.r.}) = \sigma^2 p_r q_a / n_{ar}$

où n_{ar} est le nombre d'animaux de race r alimentés avec a . Il y a orthogonalité si d'une part tous les effets sont estimables, c'est à dire si $n_{ar} > 0$ pour tout a et tout r , d'autre part les covariances entre 2 effets de nature différente sont nulles. Cette dernière condition implique en particulier, si $\hat{\mu}_{.r.} = \sum_a p_r q_a \bar{y}_{ar}$, que :

$$\text{cov}(\hat{\mu}_{a..} - \hat{\mu}_{a'..}, \hat{\mu}_{.r.}) = 0 \quad , \quad \text{cov}(\hat{\mu}_{a..} - \hat{\mu}_{a'..}, \hat{\mu}_{.r.} - \hat{\mu}_{.r'.}) = 0$$

d'où l'on déduit :

$$\text{cov}(\hat{\mu}_{a..} - \hat{\mu}_{a'..}, \hat{\mu}_{.r.}) = \sigma^2 p_r \left[\frac{q_a}{n_{ar}} - \frac{q_{a'}}{n_{a'r}} \right] = 0$$

Il existe alors une constante λ_r telle que $n_{ar} = \lambda_r q_a$ pour tout a , et de même une constante η_a telle que $n_{ar} = p_r \eta_a$. On a donc $\lambda_r / p_r = \eta_a / q_a$. Ce rapport est constant : $\lambda_r = \lambda p_r$, et finalement :

$$n_{ar} = \lambda q_a p_r \tag{4.1}$$

Il est facile de montrer que cette condition nécessaire d'orthogonalité est aussi suffisante.

Lorsqu'il y a orthogonalité, l'étude des effets principaux, de l'interaction et de la moyenne générale définis à partir de poids q_a et p_r donne l'essentiel de l'information contenue dans les données. Lorsqu'il y a un écart important à l'orthogonalité, il convient d'examiner aussi d'autres combinaisons linéaires des paramètres que l'on peut obtenir comme dans le cas de la régression multiple en

estimant dans des sous-modèles du modèle (1.8). Les effets de la non orthogonalité sont particulièrement flagrants lorsque certains effectifs n_{ar} sont nuls. Imaginons par exemple qu'il y a $k=3$ aliments, $p=2$ races et que les effectifs n_{ar} sont égaux à 2 sauf n_{31} qui est nul.

Tableau 4.1

		a		
		1	2	3
effectifs n_{ar} par aliment et race	r	2	2	0
	2	2	2	2

Les poids choisis sont $q_1=q_2=q_3=1/3$, $p_1=p_2=1/2$. On constate alors que :

- un seul contraste d'interaction est estimable : $\mu_{22}-\mu_{21}-\mu_{12}+\mu_{11}$
- l'effet race n'est pas estimable
- un seul degré de liberté de l'effet aliment est estimable : $\mu_{2.}-\mu_{1.}$

Si on estime l'effet race $\beta=\mu_{.2}-\mu_{.1}$ dans le cadre du modèle additif (1.9), on trouve :

$$\hat{\beta} = (y_{22.}-y_{21.})/2 + (y_{12.}-y_{11.})/2$$

quantité qui a pour espérance dans le modèle interactif (1.8) le contraste $(\mu_{22}+\mu_{12})-(\mu_{21}+\mu_{11})$ et s'interprète donc comme l'effet moyen de la race pour les niveaux 1 et 2 du facteur aliment.

Si dans le tableau 4.1, les 2 sont remplacés par des 10, et le 0 par un 1, on trouve les estimations suivantes de β sous les modèles interactif ($\hat{\beta}$) et additif ($\tilde{\beta}$) :

$$\hat{\beta} = ((1/3)y_{12.}+(1/3)y_{22.}+(1/3)y_{32.}) - ((1/3)y_{11.}+(1/3)y_{21.}+(1/3)y_{31.})$$

$$\tilde{\beta} = (0.46 y_{12.}+0.46 y_{22.}+0.08 y_{32.}) - (0.46 y_{11.}+0.46 y_{21.}+0.08 y_{31.})$$

Les poids 0.46, 0.46, 0.08 sont par définition ceux qui confèrent une variance minimum à $\tilde{\beta}$. Les écarts-types correspondants sont respectivement :

$$\text{écart-type } (\beta) = 0.41 \sigma$$

$$\text{écart-type } (\tilde{\beta}) = 0.30 \sigma$$

Le passage de l'estimateur β du modèle interactif à l'estimateur $\tilde{\beta}$ du modèle additif entraîne donc un gain en écart-type, qui peut être suffisant dans certaines situations pour déceler un effet race qui n'apparaissait pas dans le modèle initial.

Cet exemple montre que l'absence totale de certains traitements, c'est à dire de certaines combinaisons de niveaux des facteurs, a des conséquences beaucoup plus néfastes qu'un simple déséquilibre dans les effectifs. Pour ces cas où il y a absence de certains traitements, donc non estimabilité de certains des effets que l'on désire a priori étudier, il existe une technique élémentaire permettant de trouver pour chaque effet non estimable une combinaison linéaire simple d'effets à lui rajouter pour obtenir un effet estimable. Nous décrivons cette technique dans le cas général.

5. Recherche de fonctions estimables

Nous partirons d'une paramétrisation du type (3.24) où les paramètres $\theta_j = [z_j, \varphi]$ sont directement interprétables et non redondants. Le système d'équations correspondant est écrit sous la forme matricielle usuelle :

$$y = X \theta + \varepsilon \quad \theta = (\theta_1, \dots, \theta_p) \quad (4.2)$$

Soit $(X'X)^-$ une inverse généralisée de $X'X$, qui vérifie par définition :

$$(X'X) (X'X)^- (X'X) = (X'X) \quad (4.3)$$

L'ensemble des vecteurs a tels que $a\theta$ soit estimable coïncide avec l'image de X , qui est également l'image de $(X'X)(X'X)^-$. On peut donc prendre les colonnes de cette dernière matrice comme système générateur des fonctions estimables. Remarquons que si $X'X$ est inversible, donc si tous les θ_j sont estimables, $(X'X)^- = (X'X)^{-1}$, donc $(X'X)(X'X)^-$ est l'identité et on retrouve $\theta_1, \dots, \theta_p$ comme système

générateur des fonctions estimables. Dans tous les cas de figure, si l'un des paramètres θ_j est estimable, le vecteur canonique e_j de R^p dont les coordonnées sont nulles à l'exception de la j ème égale à 1 appartient à l'image de $'X$ (puisque $'e_j\theta = \theta_j$ est estimable) et est donc invariant par $('XX)('XX)^-$. La j ème colonne de cette dernière matrice est dans ce cas égale à e_j , et on retrouve θ_j dans le système générateur.

On notera aussi qu'à partir de toute forme linéaire $'a\theta$, on obtient une forme linéaire estimable en prémultipliant a par $('XX)('XX)^-$, à savoir la forme $'a('XX)^-('XX)\theta$. Cette forme linéaire coïncide avec $'a\theta$ si $'a\theta$ est estimable.

On obtient une inverse généralisée assez simplement en mettant $('XX)$ sous la forme :

$$('XX) = 'A D A \quad (4.3)$$

où : A est une matrice $p \times p$ inversible

D est une matrice diagonale comportant des 0 et des 1 sur la diagonale

D est une inverse généralisée d'elle même. Il s'en suit que $A^{-1}D 'A^{-1}$ est une inverse généralisée de $'XX$. Le système générateur des fonctions estimables est alors formé par les colonnes de :

$$'ADA A^{-1}D 'A^{-1} = 'AD 'A^{-1} \quad (4.4)$$

Une petite modification de la technique de triangularisation de Choleski permet d'obtenir dans (4.3) une matrice A triangulaire supérieure. On peut alors montrer que les générateurs des fonctions estimables obtenus avec une telle matrice sont identiques à ceux qu'on obtient en remplaçant progressivement dans l'expression $X\theta$ de l'espérance de y , chaque colonne de X qui est combinaison linéaire des colonnes à sa gauche par cette combinaison linéaire. Dans ce processus, on est amené à ajouter le paramètre associé à la colonne remplacée, multiplié par le coefficient ad-hoc, à chacun des paramètres des colonnes de la combinaison linéaire.

Exemple : nous reprenons l'exemple des porcs (1.8) avec les effectifs définis par le tableau 4.1 et les poids $q_1=q_2=q_3=1/3$, $p_1=p_2=1/2$. On pose :

$$R_{10} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad R_{11} = \sqrt{3/2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad R_{12} = \sqrt{1/2} \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

$$R_{20} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad R_{21} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

La base orthonormée (z_j) utilisée pour reparamétriser comme en (3.24) est définie par les 6 produits tensoriels $R_{1i} \otimes R_{2j}$. La matrice $X=(x_1, x_2, x_3, x_4, x_5, x_6)$ est alors obtenue par duplication de la matrice suivante :

		A	R	θ_1	aliment θ_2 θ_3		effets race θ_4	interaction θ_5 θ_6	
2 fois	}	1	1	1	1	-1	1	1	-1
		2	1	1	-1	-1	1	-1	-1
		1	2	1	1	-1	-1	-1	1
		2	2	1	-1	-1	-1	1	1
		3	2	1	0	2	-1	0	-2
coef. multiplicateur de la colonne				x_1	$\sqrt{3/2}$ x_2	$\sqrt{1/2}$ x_3	x_4	$\sqrt{3/2}$ x_5	$\sqrt{1/2}$ x_6

La dernière colonne x_6 est combinaison linéaire des colonnes 1, 3, 4 :

$$\sqrt{2} x_6 = -x_1 - \sqrt{2} x_3 - x_4$$

On a donc :

$$E(y) = \theta_1 x_1 + \dots + \theta_6 x_6 = \theta_1 x_1 + \dots + \theta_6 (-x_1/\sqrt{2} - x_3 - x_4/\sqrt{2}) =$$

$$= (\theta_1 - \theta_6/\sqrt{2})x_1 + \theta_2 x_2 + (\theta_3 - \theta_6)x_3 + (\theta_4 - \theta_6/\sqrt{2})x_4 + \theta_5 x_5$$

Les vecteurs x_1, x_2, x_3, x_4, x_5 engendrent un espace de dimension 5 et sont indépendants. Il en découle que les combinaisons linéaires :

$$(\theta_1 - \theta_6/\sqrt{2}) \quad \theta_2 \quad (\theta_3 - \theta_6) \quad (\theta_4 - \theta_6/\sqrt{2}) \quad \theta_5$$

forment un système générateur des fonctions estimables. Compte tenu des identités :

$$\theta_3 = \sqrt{1/2} [-\mu_{11} - \mu_{21} + 2\mu_{31} - \mu_{12} - \mu_{22} + 2\mu_{32}]$$

$$\theta_4 = \mu_{11} + \mu_{21} + \mu_{31} - \mu_{12} - \mu_{22} - \mu_{32}$$

$$\theta_6 = \sqrt{1/2} [-\mu_{11} - \mu_{21} + 2\mu_{31} + \mu_{12} + \mu_{22} - 2\mu_{32}]$$

$$\begin{aligned} \text{on a : } \theta_3 - \theta_6 &= \sqrt{2} [2\mu_{32} - \mu_{12} - \mu_{22}] \\ \theta_4 - \theta_6 / \sqrt{2} &= (3/2) [(\mu_{11} + \mu_{21}) - (\mu_{12} + \mu_{22})] \end{aligned}$$

Ces 2 dernières combinaisons linéaires s'interprètent donc simplement comme un effet aliment pour la race 2 d'une part, un effet race moyen pour les aliments 1 et 2 d'autre part. On peut aussi dire que les estimateurs correspondants sont des estimateurs de θ_3 et θ_4 respectivement, qui sont biaisés par un terme d'interaction, lequel terme est ici totalement inestimable.

ANNEXE : polynomes orthogonaux dans $L_2(\mu)$ dans le cas où la probabilité μ est symétrique autour de 0.

$$\text{On note } m_s \text{ le moment d'ordre } s \text{ de } \mu : m_s = \int x^s d\mu(x)$$

Les 4 premiers polynomes orthogonaux ayant 1 pour coefficient du monome de degré le plus élevé sont donnés ci-dessous avec le carré de leur norme.

$$\begin{aligned} P_0 &= 1 & [P_0, P_0] &= m_0 = 1 \\ P_1 &= x & [P_1, P_1] &= m_2 \\ P_2 &= x^2 - \frac{m_2}{m_0} & [P_2, P_2] &= m_4 - \frac{m_2 m_2}{m_0} \\ P_3 &= x^3 - \frac{m_4}{m_2} x & [P_3, P_3] &= m_6 - \frac{m_4 m_4}{m_2} \\ P_4 &= x^4 - \frac{m_6 m_0 - m_4 m_2}{m_4 m_0 - m_2 m_2} x^2 + \frac{m_6 m_2 - m_4 m_4}{m_4 m_0 - m_2 m_2} & [P_4, P_4] &= m_8 - m_6 \frac{m_6 m_0 - m_4 m_2}{m_4 m_0 - m_2 m_2} - m_4 \frac{m_6 m_2 - m_4 m_4}{m_4 m_0 - m_2 m_2} \end{aligned}$$

Le polynome normé R_j se déduit immédiatement de P_j par :

$$R_j = P_j / \sqrt{[P_j, P_j]}$$

BIBLIOGRAPHIE

- BAILEY R.A., PRAEGER C.E., ROWLEY C.A., SPEED T.P.**(1983). Generalized wreath product of permutation groups. Proc. London Math. Soc. **47** 69-82.
- DENIS J.B.**(1988). Two Way Analysis Using Covariates. Statistics **19** 1, 123-132
- DENIS J.B.**(1980). Analyse de régression factorielle. Biométrie-Praximétrie, **20**,1-34.
- KOBILINSKY** (1985). Orthogonal factorial designs for quantitative factors. Statistics and Decisions, Supplement 2 272-285.
- MODULAD** (1987). Bibliothèque FORTRAN 77 pour l'analyse des données. Module MODLI p361-395. INRIA.
- RAKTOE B.L., HEDAYAT A., FEDERER W.T.**(1981) Factorial designs. Wiley. New-York.
- SAS** (1985). SAS User's Guide Statistics. Version 5 Edition. The four types of estimable functions p77-92. SAS Institute Inc.
- TJUR T.** (1984). Analysis of variance models in orthogonal designs (with discussion). Internat. Statist. Review **52** 33-65.