PRÉSENTATION GÉNÉRALE
DU PROGRAMME PAC

PLAN D'ANALYSE

LA STRUCTURE (SUJETS, GROUPES, OCCASIONS)

La classe des plans d'analyse traités par **PAC** est formalisée par une structure (*Sujets, Groupes, Occasions*), qui est caractérisée de la manière suivante.

▶ Sujets

Le plan comporte un seul facteur aléatoire, désigné par le terme Sujets. Ce facteur pourra correspondre effectivement à des sujets mais recevra de manière générale une acception abstraite.

▶ Groupes

Les Groupes sont un ensemble dans lequel le facteur Sujets est emboîté : chaque sujet s est affecté à un seul groupe g. Les effectifs des différents groupes peuvent être ou non égaux (plan équilibré ou non-équilibré) ; dans le second cas on a le choix, pour chaque comparaison, entre une procédure "équipondérée" et une procédure "pondérée par les effectifs".

Les Groupes constituent donc, suivant les différentes terminologies en usage, un facteur systématique (*fixed-effects*) de type *grouping*, *between-groups*, *whole-plot*...

▶ Occasions

Comme cas privilégié, les Occasions peuvent être un ensemble croisé avec le facteur Groupes : on a dans ce cas exactement une observation pour chaque combinaison possible des sujets et des occasions. Les Occasions constituent alors, suivant les différentes terminologies en usage, un facteur systématique de type *within*, *repeated-measures*, *trial*, *split-plot*...

Mais la relation entre les Groupes et les Occasions peut parfaitement être quelconque : par exemple, les sujets d'un groupe g n'ont d'observations que pour une partie des occasions ; ceci permet de traiter des plans incomplets.

Cette classe de plans (Sujets, Groupes, Occasions) est en fait d'une très grande généralité, puisque chacun des ensembles Groupes et Occasions peut être décrit par des facteurs systématiques eux-mêmes en relation quelconque.

La seule restriction, en dehors de la prise en considération d'un seul facteur aléatoire, est que, à l'intérieur d'un même groupe, les sujets doivent avoir des observations pour les mêmes occasions.

EXEMPLE

A titre d'illustration, pour nous en tenir aux relations usuelles, nous considérerons un plan comportant :

- quatre groupes définis par le croisement d'un facteur Conditions (expérimentales) à deux modalités et du facteur Sexe ;
- huit occasions consistant en huit essais emboîtés dans deux blocs successifs de quatre essais.

C'est en fonction de la structure générale (Sujets, Groupes, Occasions) que sont déterminées les procédures inférentielles appropriées. Pour chaque comparaison analysée, les structures particulières des facteurs décrivant les Groupes et les Occasions sont automatiquement prises en compte dans la dérivation des données pertinentes. En particulier la validité de chaque demande est vérifiée : ainsi, dans l'exemple considéré, la demande d'une comparaison d'interaction sera acceptée entre les facteurs Conditions et Blocs, mais sera rejetée entre les Essais et les Blocs.

LA DESCRIPTION DU PLAN D'ANALYSE

D'une manière générale, chaque facteur élémentaire déclaré dans le plan d'analyse sera désigné par la première lettre de son nom. Les modalités (niveaux) de chaque facteur seront représentées par la lettre correspondante indiciée à l'aide des premiers entiers. Par exemple les deux modalités du facteur C (pour Conditions) seront représentées par C1 et C2. Pour faciliter la distinction entre facteurs et modalités, il est possible (et judicieux) d'utiliser respectivement des lettres majuscules (C) et des lettres minuscules (c1 et c2).

Les modalités des facteurs élémentaires définissant les Groupes doivent figurer dans le fichier des données (ou être construites par transformation). La liste des modalités des facteurs associées à chaque occasion est fournie par l'utilisateur. Ceci permet de prendre en compte des relations quelconques entre ces facteurs.

Dans l'exemple considéré ici, on pourra ainsi déclarer les facteurs :

- pour les groupes,
  - Conditions C = {c1,c2}
  - Sexe S = {s1,s2}
- pour les occasions,
  - Essais E = {e1,e2,e3,e4,e5,e6,e7,e8}
  - Blocs B = {b1,b2}

La structure complète associée à ces facteurs systématiques est représentée dans le tableau ci-après, où "xx" indique qu'il y a des observations pour la combinaison des modalités g et o correspondantes. Nous appellerons ce tableau le tableau G∗O ; ici le symbole "∗" désigne simplement le produit cartésien de G (Groupes) et O (Occasions) et ne signifie pas que G et O sont nécessairement croisés.

	o1 = e1 b1	o2 = e2 b1	o3 = e3 b1	o4 = e4 b1	o5 = e5 b2	o6 = e6 b2	o7 = e7 b2	o8 = e8 b2
g1=c1s1	xx	xx	xx	xx	xx	xx	xx	xx
g2=c2s1	xx	xx	xx	xx	xx	xx	xx	xx
g3=c1s2	xx	xx	xx	xx	xx	xx	xx	xx
g4=c2s2	xx	xx	xx	xx	xx	xx	xx	xx

(dans ce cas particulier Groupes et Occasions sont croisés)

PAC examine et vérifie les informations contenues dans le fichier des données et les instructions fournies sur les facteurs élémentaires, les variables et covariables à utiliser, etc. Il établit la structure des groupes et des occasions ; puis il calcule pour chaque combinaison groupe & occasion présente dans le plan d'analyse les statistiques de base suivantes : nombre d'observations (c'est-à-dire l'effectif du groupe en jeu), moyenne, écart-type (corrigé), minimum, maximum.

UN APERCU DU LANGAGE DES COMPARAISONS

Nous esquisserons ici un aperçu des possibilités du langage des comparaisons. Chaque formule utilise les lettres désignant les facteurs systématiques déclarés par l'utilisateur et un certain nombre de symboles, entièrement hiérarchisés, dont l'usage est facile à apprendre.

EXEMPLES

▶ Obtenir la décomposition canonique usuelle

Par exemple, à partir des facteurs systématiques élémentaires C, S, E, B, on peut obtenir la décomposition canonique usuelle par l'ensemble des demandes :

"C" "S" "B"
comparaisons globales à un degré de liberté (*dl*)

"E(B"
comparaison E intra B
ou somme des comparaisons partielles restreintes à b1 et b2, à 6 *dl*

"C.S" "C.B" "S.B" "C.E B" "S.E(B" "C.S.B" "C.S.E(B"
interactions entre les comparaisons précédentes

▶ Affiner la décomposition canonique

On peut aisément affiner cette décomposition, à partir des possibilités suivantes.

• On peut préciser par l'une des demandes

"c1,c2" ou "c2,c1" (au lieu de "C")

dans quel sens définir la différence entre les deux moyennes à comparer.

• On peut analyser
- les comparaisons intra partielles de E dans B, par l'une des demandes équivalentes

"e1,e2,e3,e4" ou "E(b1" ou "E/b1" ou ...
"e5,e6,e7,e8" ou "E(b2" ou "E/b2" ou ...

- d'une manière générale n'importe quelle sous-comparaison portant sur des parties de E

"e4,e1"
"e1_e2,e3_e4" ("_" est l'opérateur de *réunion*)

- leurs interactions par exemple avec C

"C.e1,e2,e3,e4" ou "C.E(b1" ou "c1,c2.E/b1" ou ...

etc.

• On peut étudier par exemple
- la composante linéaire de la régression polynômiale sur E tout entier ou partiellement (avec la possibilité de définir des intervalles inégaux entre les modalités)

"LIN E"
"LIN e1,e2,e3,e4" ou "LIN E/b1" ou ...

- leurs résiduelles

"-LIN E"
"-LIN e1,e2,e3,e4" ou "-LIN E/b1" ou ...

- et les interactions

"C.LIN E"
"c1,c2.-LIN E/b1"

▶ **Choisir une décomposition mieux adaptée**

Mais on peut également choisir une autre décomposition qui serait mieux adaptée aux questions posées, et par exemple, au lieu des interactions, analyser les effets du facteur C restreints à chaque sexe et chaque bloc

"c1,c2/s1b1" "c1,c2/s1b2" "c1,c2/s2b1" "c1,c2/s2b2"

Si l'objectif est de montrer que chacun de ces quatre effets partiels est négligeable, on peut analyser leur "somme" (comparaison à 4 degrés de liberté) par la demande :

"c1,c2(s1b1,s1b21,s2b1,s2b2" ou "c1,c2(S∗B"

Etc.

INTERPRÉTATION D'UNE FORMULE

L'interprétation des formules repose uniquement sur des règles ensemblistes, faciles à expliciter. Ceci confère au langage à la fois une grande intelligibilité et une souplesse considérable.

▶ **Partition du tableau G∗O**

Dans une première étape est construite une partition du tableau G∗ O. Cette étape peut être contrôlée par l'utilisateur, puisque les classes (parties de G∗ O associées à chaque formule valide) et les moyennes de ces classes pour les différentes variables sont fournies en sorties.

▶ **Dérivations sur les moyennes**

Dans une seconde étape, la formule de comparaison reçoit une interprétation algébrique, en termes de dérivations sur les moyennes. **PAC** engendre automatiquement les combinaisons linéaires appropriées, qui seront appliquées aux moyennes du tableau G∗ O pour définir l'effet de la comparaison. Ici encore l'utilisateur peut obtenir la sortie de ces combinaisons.

COMBINAISON LINÉAIRE ASSOCIÉE À UNE COMPARAISON À UN DEGRÉ DE LIBERTÉ

A l'issue de la première étape, des coefficients numériques sont affectés, de manière canonique et récursive, aux classes de la partition. Les combinaisons linéaires des moyennes associées à ces coefficients définissent le *représentant de l'effet* de la comparaison. Une comparaison à un degré de liberté est représentée par un seul vecteur de coefficients ; et une comparaison à m degrés de liberté par une famille de m vecteurs.

EXEMPLES

Cette procédure est brièvement illustrée dans les exemples suivants de comparaisons à un degré de liberté, qui constituent les situations de base.

▶ **Inférence sur la moyenne d'une partie**

La formule "c1b1" engendre la partie de G∗ O

$$\{g1,g2\} \times \{o1,o2,o3,o4\}$$

L'effet de la comparaison est la moyenne de cette partie, qui est donc affectée du coefficient 1.

▶ Comparaison des moyennes de deux parties

La formule "c1,c2/b1" engendre la partition de G∗O en deux classes (associées respectivement à "c1b1" et à "c2b1")

$$\{g1,g2\} \times \{o1,o2,o3,o4\} \text{ et } \{g3,g4\} \times \{o1,o2,o3,o4\}$$

L'effet de la comparaison est la différence des moyennes de ces deux parties, qui sont donc affectées respectivement des coefficients +1 et -1 (contraste privilégié).

▶ Interaction de comparaisons des moyennes de deux parties

La formule "c1,c2.b1,b2" (interaction entre les comparaisons "c1,c2" et "b1,b2") engendre la partition obtenue par le croisement des partitions en deux classes associées respectivement à "c1,c2" et à "b1,b2", d'où les quatre classes (associées respectivement à "c1b1", "c1b2", "c2b1" et "c2b2")

$$\begin{array}{ll} \{g1,g2\}\times\{o1,o2,o3,o4\} & \{g3,g4\}\times\{o1,o2,o3,o4\} \\ \{g1,g2\}\times\{o5,o6,o7,o8\} & \{g3,g4\}\times\{o5,o6,o7,o8\} \end{array}$$

Ces quatre parties seront affectées des coefficients $+1, -1, -1, +1$, produit terme à terme des contrastes privilégiés affectés respectivement à "c1,c2" et à "b1,b2", d'où la caractérisation habituelle de l'effet d'interaction comme différence de différences.

Ensuite les coefficients ainsi définis sont remontés sur G∗O, c'est-à-dire, pour chaque partie, "répartis" sur les modalités (g,o) qui la composent. Pour un plan non-équilibré cette répartition, selon l'option choisie par l'utilisateur, est faite, soit de façon uniforme, soit proportionnellement aux effectifs associés à chaque modalité.

COMPARAISON A PLUSIEURS DEGRÉS DE LIBERTÉ

Une comparaison à plusieurs degrés de liberté est décomposée en comparaisons à un degré de liberté de manière à se ramener aux situations de base.

Par exemple, la comparaison "e1,e2,e3,e4" consiste à comparer globalement les moyennes de quatre parties.Elle est décomposée en trois sous-comparaisons entre deux parties et une base orthogonale de coefficients appropriés est choisie, de manière à caractériser la grandeur de l'effet de la comparaison comme la moyenne quadratique (pondérée s'il y a lieu) des différences entre les moyennes des quatre parties prises deux à deux.

COMPARAISON DE TYPE PRODUIT

Toutes les comparaisons qui ont été prises en exemple présentent une propriété remarquable relativement à la structure (Groupes, Occasions) : les coefficients sur G∗O qui leur sont associées peuvent être obtenus en effectuant le produit terme à terme de coefficients marginaux sur g et de coefficients marginaux sur O. Une telle comparaison sera dite "une comparaison de type produit" par rapport à la structure (Groupes, Occasions).

Cette propriété est illustrée ci-après pour les comparaisons "c1,c2/b1" et "c1,c2.b1,b2" (nous avons considéré ici la solution équipondérée).

"c1,c2/b1"

	o1 = e1 b1	o2 = e2 b1	o3 = e3 b1	o4 = e4 b1	o5 = e5 b2	o6 = e6 b2	o7 = e7 b2	o8 = e8 b2	
g1=c1s1	+1/8	+1/8	+1/8	+1/8	0	0	0	0	+1/2
g2=c2s1	−1/8	−1/8	−1/8	−1/8	0	0	0	0	−1/2
g3=c1s2	+1/8	+1/8	+1/8	+1/8	0	0	0	0	+1/2
g4=c2s2	−1/8	−1/8	−1/8	−1/8	0	0	0	0	−1/2
	1/4	1/4	1/4	1/4	0	0	0	0	

“c1,c2.b1,b2”

	o1	o2	o3	o4	o5	o6	o7	o8	
	=	=	=	=	=	=	=	=	
	e1	e2	e3	e4	e5	e6	e7	e8	
	b1	b1	b1	b1	b2	b2	b2	b2	
g1=c1s1	+1/8	+1/8	+1/8	+1/8	-1/8	-1/8	-1/8	-1/8	+1/2
g2=c2s1	−1/8	−1/8	−1/8	−1/8	+1/8	+1/8	+1/8	+1/8	−1/2
g3=c1s2	+1/8	+1/8	+1/8	+1/8	-1/8	-1/8	-1/8	-1/8	+1/2
g4=c2s2	−1/8	−1/8	−1/8	−1/8	+1/8	+1/8	+1/8	+1/8	−1/2
	+1/4	+1/4	+1/4	+1/4	-1/4	-1/4	-1/4	-1/4	

Les coefficients sur G sont dans les deux exemples le contraste sur G [+1/2 -1/2 +1/2 -1/2] qui oppose les groupes g1 et g3 aux groupes g2 et g4.

Les coefficients sur O correspondent dans le premier cas au moyennage sur les quatre occasions associées à la modalité b1 et dans le second cas au contraste qui oppose ces quatre occasions aux quatre restantes.

▶ Trois catégories de comparaisons de type produit

Les comparaisons de type produit sont classées par **PAC** en trois catégories principales :
- Comparaisons entre les Groupes restreintes à une partie des Occasions (éventuellement la partie pleine) ; par exemple

“c1,c2/b1” “S” “C.S/e1” ...

- Comparaisons entre les Occasions restreintes à une partie des Groupes (éventuellement la partie pleine) ; par exemple

“b1,b2/c1” “E(B” “E/b2c2” ...

- Interactions entre les deux types précédents ; par exemple

“b1,b2.c1,c2” “E(B.c1,c2” “C.S.E/b1” ...

▶ Des procédures générales de calcul

Les comparaisons de type produit constituent la grande majorité des comparaisons usuelles ; elles sont privilégiées, dans la mesure où les suppositions habituelles sur les variances et covariances (homogénéité, éventuellement circularité) peuvent être raisonnablement envisagées (sinon satisfaites). Elles correspondent dans **PAC** aux mêmes procédures générales de calcul.

▶ Les comparaisons qui ne sont pas de type produit

En outre des procédures particulières permettent d'analyser des comparaisons qui ne sont pas de type produit (dans la version actuelle de **PAC** cette possibilité est toutefois limitée aux comparaisons à un degré de liberté).

ANALYSE SPÉCIFIQUE D'UNE COMPARAISON

Chaque comparaison donne lieu à une analyse *spécifique*.

DÉRIVATION DES DONNÉES PERTINENTES

▶ Coefficients de dérivation sur les Groupes et sur les Occasions

Pour une comparaison de type produit, deux bases marginales orthogonales de combinaisons linéaires sont construites respectivement sur O et sur G (voir les exemples précédents). Ces bases fournissent respectivement les "coefficients de dérivation sur les Occasions" et les "coefficients de dérivation sur les Groupes". L'utilisateur peut obtenir ces coefficients dans les sorties, et contrôler ainsi tous les calculs.

▶ Caractérisation de la grandeur l'effet

La formalisation algébrique de l'Analyse des Comparaisons fournit des algorithmes simples et généraux (en particulier indépendants des relations entre les facteurs systématiques du plan) pour construire une base de la comparaison. Par cette formalisation on munit l'espace vectoriel des observations associé à toute dérivation linéaire d'une métrique appropriée, qui permet à la fois de prendre en compte le cas des effectifs inégaux et de retrouver les sommes-des-carrés-et-produits exactes directement à partir des données pertinentes.

Pour une comparaison à un degré de liberté l'effet est de manière naturelle une combinaison linéaire de moyennes : une différence de moyennes, une différence de différences de moyennes (interaction), la pente d'une droite de régression,... Nous noterons δ l'effet *vrai* (calculé à partir des moyennes parentes).

Les comparaisons à plusieurs degrés de liberté sont décomposées en comparaisons à un degré de liberté. Par exemple, la comparaison globale de trois moyennes est décomposée en deux sous-comparaisons de deux moyennes. Une base orthogonale de coefficients peut être choisie de manière à caractériser la grandeur de l'effet de la comparaison comme la moyenne quadratique (pondérée s'il y a lieu) des différences entre les moyennes prises deux à deux. Nous noterons λ la grandeur de l'effet vrai ; en particulier l'hypothèse nulle usuelle d'absence d'effet s'écrit $\lambda = 0$.

▶ Protocole dérivé pertinent

Le protocole dérivé des données pertinentes pour la comparaison est construit en calculant pour chaque sujet les valeurs associées aux différentes combinaisons linéaires de la base sur les Occasions.

Ces données pertinentes sont d'une interprétation claire ; par exemple, pour la comparaison "c1,c2/b1" elles sont constituées, pour chaque sujet, de la moyenne de ses observations pour l'ensemble des essais du bloc 1. Pour la comparaison "c1,c2.b1,b2" elles sont constituées, pour chaque sujet, de la différence entre la moyenne des essais du bloc 1 et celle des essais du bloc 2 (effet observé "individuel" de la comparaison "b1,b2"). En sortie on peut obtenir pour chaque groupe les moyennes des données pertinentes, les écarts-types et les corrélations (entre les différentes combinaisons linéaires et entre les variables).

Les données pertinentes individuelles peuvent en outre être sauvegardées sur un fichier pour d'autres analyses.

MODÈLES SPÉCIFIQUES ET CONDITIONS DE VALIDITÉ

La dérivation sur les Occasions permet dans chaque cas de se ramener à une structure de groupes indépendants, avec des données univariées ou multivariées, et d'appliquer les solutions standard pour ce type de plans. On accroît ainsi de manière décisive la simplicité et le réalisme des procédures ; en particulier les conditions de validité peuvent être facilement explicitées et être rendues très faibles.

PAC fournit des analyses variable par variable et des analyses sur l'ensemble des variables et propose dans chaque cas, différentes solutions, suivant qu'elles mettent ou non en jeu des suppositions sur les variances et covariances du modèle. Ces procédures englobent les approches traditionnelles, "univariées" et "multivariées" des plans à mesures répétées (y compris les solutions fréquentistes approchées de correction).

Grâce à l'approche de l'analyse spécifique, les procédures inférentielles sont construites directement à partir de statistiques descriptives élémentaires, faciles à interpréter, et en tout premier lieu de l'effet observé (éventuellement multivarié).

En particulier pour toute comparaison à un degré de liberté, les tests de signification univariés sont présentés explicitement, sous la forme d'un t de Student. Celui-ci est construit comme le rapport de l'effet observé (une combinaison linéaire de moyennes) à un indice d'échelle, défini comme un écart-type moyen divisé par une quantité homogène à la racine-carrée d'un effectif. Le carré de ce t n'est autre que le rapport F usuel des carrés-moyens.

Le même type d'analyse est généralisé aux comparaisons à plusieurs degrés de liberté et aux procédures multivariées

DES PROCÉDURES INFÉRENTIELLES POUR RECHERCHER DES CONCLUSIONS SUR L'IMPORTANCE DES EFFETS

PROCÉDURES FRÉQUENTISTES

► Test d'une hypothèse ponctuelle quelconque (un degré de liberté)

Pour toute comparaison à un degré de liberté l'effet vrai δ (univarié ou multivarié) peut être comparé à n'importe quelle valeur δ_0 : "hypothèse nulle ponctuelle", $\delta = \delta_0$.

► Test d'une hypothèse composée

Pour chaque variable, on peut tester l'hypothèse que la grandeur de l'effet vrai est supérieure (ou est inférieure) à une limite donnée x : "hypothèse nulle composée", par exemple pour un degré de liberté $\mid \delta \mid > x$ (ou $\mid \delta \mid < x$).

► Intervalles de confiance fréquentiste

- Pour toute comparaison à un degré de liberté, on obtient l'intervalle de confiance fréquentiste usuel pour l'effet de chaque variable : intervalle centré sur l'effet observé.

- Pour toute comparaison, on peut également obtenir un intervalle de confiance pour la grandeur de l'effet $\lambda(\mid \delta \mid$ pour un degré de liberté) de chaque variable : intervalle du type $[0, x]$ ou $[x, +\infty[$.

PROCÉDURES FIDUCIO-BAYÉSIENNES

PAC fournit, pour chaque comparaison analysée, la distribution de probabilité fiducio-bayésienne sur l'effet vrai de cette comparaison. Différents énoncés fiducio-bayésiens permettent d'obtenir directement des conclusions sur l'importance des effets examinés.

Ces énoncés donnent :

- soit les limites correspondant à une probabilité (*garantie*) choisie par l'utilisateur, ceci à la fois pour chacune des variables et sur l'ensemble des variables ;

- soit les probabilités correspondant à une limite x fournie par l'utilisateur (pour chacune des variables).

AUTRES POSSIBILITÉS

PAC propose de nombreuses options qui conduisent dans chaque cas à des analyses explicites et détaillées. Nous mentionnerons encore ici les possibilités suivantes.

► Régression polynômiale

On peut effectuer des analyses détaillées des différentes composantes de la régression polynômiale, sur des parties affectées de coefficients quelconques.

► Prise en compte de covariables

On peut prendre en compte une ou plusieurs covariables. L'approche de l'analyse spécifique permet facilement de considérer des liaisons polynômiales entre les variables et les covariables.

► Analyse discriminante

On peut obtenir les combinaisons linéaires des variables (analyse discriminante) et/ou les contrastes d'une comparaison donnée (appelés contrastes essentiels) qui maximisent la grandeur de l'effet observé.