

même nombre de modalités, par exemple 2) peuvent être tels que :

$$[A \times B] < C > ,$$
$$[B \times C] < A > ;$$
$$[C \times A] < B > ,$$

donc être tels que tous les facteurs composés binaires sont cette fois des croisements ; mais pour que le facteur composé de A,B,C soit quasi-complet il faudrait en outre que les trois facteurs A,B,C soient croisés ce qui est impossible car on ne peut avoir à la fois (C n'étant pas constant) :

$$[A \times B] < C > \text{ et } [A \times B] \times C.$$

(N.B. : la structure précédente n'est autre que la structure, ~~très~~ courante dans l'expérimentation, de carré latin construit sur 3 facteurs ; nous rencontrerons une illustration de cette structure, au chapitre VIII, avec les "données de Cochran et Cox", avec un carré latin construit sur les trois facteurs Machines, Essais et Ordres).

CHAPITRE IV - DE L'EXPLORATION DES DONNÉES AUX ANALYSES FINES : FORMULES D'INTERROGATION ET DEMANDES D'ANALYSE

Exploration des données et examens à vue.

Dans une première phase de l'analyse des données expérimentales, que nous appellerons l'exploration des données, l'expérimentaliste construit un certain nombre de tableaux et graphiques et procède à leur examen à vue.

Cette phase, tout en étant guidée par les objectifs qui ont suscité la recherche (par exemple, pour les "données H & B" : l'examen de l'interaction entre deux facteurs expérimentaux A et B) sera essentiellement "ouverte", et l'expérimentaliste en profitera généralement pour "recouper" des résultats antérieurs (par exemple, ici : les effets de chacun de ces deux facteurs expérimentaux, pris séparément).

Du point de vue formel, tous ces examens constitueront des dérivations effectuées sur le protocole. Nous allons maintenant, sur les "données H & B", illustrer les plus simples de ces dérivations : la restriction et le moyennage.

Restriction à une partie d'un facteur ; restriction à un facteur-partiel, à une modalité d'un facteur

Supposons qu'au cours de l'exploration des données, nous voulions nous restreindre aux données d'une partie des sujets, disons les sujets s1 à s6 (*). Ces données constitueront un sous-protocole (strict) du protocole d'ensemble, dérivé par restriction à une partie du facteur S, en l'occurrence la partie (stricte) $S' = \{s1, s2, \dots, s6\}$ [Si l'on considère le Tableau I, ce sous-protocole correspondra à une partie de ce tableau].

Pour décrire ce sous-protocole, on prendra naturellement les mêmes descripteurs que pour le protocole d'ensemble ; d'où le plan $R\langle S' * A * B * C \rangle$ qui est un facteur-partiel du plan $R\langle S * A * B * C \rangle$ décrivant le protocole d'ensemble ; la formule $R\langle S' * A * B * C \rangle$ désignera également sans ambiguïté le sous-protocole correspondant.

(*) Méthodologiquement, une telle restriction ira de soi lorsque la partie du facteur considérée correspond à une modalité d'un autre facteur (par exemple, si les sujets s1 à s6, à la différence des autres, avaient été soumis à une certaine tâche préalable à l'expérimentation). Dans l'expérience rapportée ici, ce n'était pas le cas ; la sélection des sujets s1 à s6 serait donc une sélection effectuée à partir des observations (rappelons que nous avons numéroté les sujets selon l'ordre croissant de leurs effets d'interaction : cf. tableau IV). Sans pousser la discussion méthodologique, disons qu'une telle sélection a posteriori ne soulèvera pas de problème majeur dans la phase d'exploration des données, mais deviendrait plus délicate si elle était effectuée en vue de conclusions inférentielles (cf. chapitre V, la discussion des tests de signification "a posteriori").

Un cas particulier sera celui de la restriction à un sujet donné, disons s_1 ; d'où le protocole que l'on dira dérivé par restriction à la modalité s_1 du facteur S ; le plan de ce protocole sera le facteur-partiel (ou plan-partiel) $R\langle s_1 \rangle * A * B * C$, qu'on écrira simplement $R\langle s_1 \rangle * A * B * C$ (cf. les commentaires du chapitre précédent au sujet de l'écriture $F\langle g \rangle$).

Dans ce qui suit : à chacun des sujets nous associerons le sous-protocole ainsi obtenu par restriction ; chacun de ces protocoles, bien que dérivé (par rapport au protocole d'ensemble) sera qualifié de protocole individuel "fondamental", car il servira de base aux "analyses individuelles" du sujet correspondant.

Moyennage sur un facteur

Supposons maintenant que nous calculions, pour chaque sujet et chaque condition, la moyenne des observations relatives à ce sujet et à cette condition ; la famille de ces moyennes constituera un protocole dérivé par moyennage sur (les modalités d') un facteur : ici, le facteur Répétitions $R\langle \rangle$ (le rappel de la structure d'emboîtement du facteur R est destiné à souligner que chacune des opérations de moyennage porte, bien entendu, non sur toutes les modalités du facteur R, mais seulement sur celles qui correspondent à une même modalité du facteur emboîtant).

Le protocole précédent sera qualifié de "protocole de groupe fondamental", car c'est à partir de ce protocole que l'on procédera aux "analyses de groupe".

Ce protocole (lorsque on prend $C=C_2$) est reproduit dans la sortie n° 1 de l'Annexe ; il admet pour plan $S * A * B * C$, qui est un sous-facteur du plan $R\langle S * A * B * C \rangle$. La formule $S * A * B * C$ désignera sans ambiguïté ce plan, ainsi que le protocole correspondant, si l'on adopte la convention de notation suivante, dite de la notation par défaut : l'absence d'une lettre, aussi bien en majuscule qu'en minuscule (ici : l'absence de la lettre R) signifie qu'on a moyenné sur le facteur correspondant.

En procédant de la sorte à diverses restrictions et moyennages successifs, on obtiendra des protocoles de plus en plus "condensés" dont les supports seront des sous-facteurs, ou sous-facteurs-partiels. Dans le tableau II, on a présenté quelques "chaînes de dérivation" envisageables.

On remarquera qu'un même protocole dérivé peut être obtenu à partir de plusieurs chaînes de dérivation. Ainsi, le protocole $s_1 * A * B * C$ peut être obtenu à partir du protocole de groupe fondamental $S * A * B * C$, par restriction à s_1 , mais aussi bien à partir du protocole individuel fondamental $R \langle s_1 * A * B * C \rangle$, par moyennage sur le facteur $R \langle \rangle$.

Cas où le plan est un croisement : protocole dérivé conditionnel

La notion de protocole dérivé conditionnel constitue une généralisation des deux notions de protocole dérivé par restriction et par moyennage ; mais pour introduire cette notion, nous nous limiterons au cadre particulier d'un protocole admettant pour plan un croisement ; comme illustration, nous prendrons le protocole de groupe fondamental, dont le plan $S * A * B * C$ sera considéré ici comme le croisement du facteur S et du facteur $A * B * C$. A partir de ce protocole :

- on peut procéder à une restriction à une partie S' du facteur S ; d'où le sous-protocole admettant pour plan $S' * A * B * C$, plan-partiel de $S * A * B * C$ (Dans le "tableau des données" de la sortie n° 1 de l'Annexe (lorsqu'on prend $C=C_2$) : ce sous-protocole correspond à l'ensemble des lignes du tableau relatives à la partie S' ; par exemple, si $S' = \{s_1, s_2, \dots, s_6\}$, ce sous-protocole est constitué par les 6 premières lignes du tableau ;

- en particulier, on peut procéder à une restriction à une modalité (disons s_1) du facteur S ; d'où le protocole (qui sera qualifié d'"individuel") ; ce protocole admettra pour plan $s_1 * A * B * C$. Tous les plans de ces protocoles individuels seront en correspondance bijective deux à deux, ainsi qu'avec le facteur $A * B * C$, sous-facteur de $S * A * B * C$. (Dans la sortie n° 1 de l'Annexe, ces protocoles individuels correspondent aux lignes du "tableau des données", indexées par s_1, s_2, \dots).

- on peut procéder à un moyennage sur le facteur S : d'où le protocole dérivé (qu'on qualifiera "de groupe") admettant pour plan le facteur $A * B * C$, sous-facteur de $S * A * B * C$; l'écriture $A * B * C$ désignera sans ambiguïté, moyennant la convention par défaut, le plan et le protocole dérivé lui-même. (Dans la sortie n° 1 de l'Annexe : ce protocole de groupe correspond au "tableau des moyennes").

On pourrait également envisager de moyennner sur une partie S' du facteur S. Cette dérivation, clairement, contiendra comme cas particuliers, d'une part la restriction à une modalité de S (en prenant $S' = \{s_1\}$ par exemple), d'autre part le moyennage sur le facteur S (en prenant $S'=S$). Mais lorsque S' est une partie, stricte et non-réduite à 1 élément de S, il s'agira d'une dérivation nouvelle, que nous appellerons moyennage sur le facteur partiel S'. Dans le cas général, nous dirons que le protocole dérivé est un protocole dérivé conditionnel (le conditionnement portant sur la partie S' de S).

Plan d'un protocole conditionnel ; symbole de conditionnement et notation pleine pour le moyennage

Reprenons l'exemple du protocole conditionnel, dérivé, à partir du protocole $S*A*B*C$, par moyennage sur une partie S' (quelconque) de S. Le support de ce protocole est en correspondance bijective avec $A*B*C$, mais l'écriture $A*B*C$, qui ne fait pas apparaître S', est insuffisante pour caractériser le protocole. Pour écrire le plan de ce protocole conditionnel, nous adopterons la formule suivante :

$$A*B*C/S'$$

Cette écriture met en jeu le nouveau symbole "/" qu'on appellera barre de conditionnement, ou barre transversale, ou "slash". Dans cette écriture, la partie du facteur sur laquelle on a moyenné apparaîtra donc à droite du symbole "/" ; par exemple, si $S' = s_1, s_2, \dots, s_6$, on écrirait :

$$A*B*C/s_1 s_2 s_3 s_4 s_5 s_6$$

[On pourra ici considérer que l'écriture avec le symbole de conditionnement est, par définition, équivalente à l'écriture ensembliste $\{S'\}*A*B*C$, où $\{S'\}$ désigne le facteur (constant) ayant pour unique modalité la partie S']

Dans le cas particulier d'une restriction à une modalité, disons s_1 , on écrirait de même la formule :

$$A*B*C/s_1$$

avec, à droite du symbole "/", la modalité (écrite en minuscules) du facteur sur lequel porte la restriction (cette formule sera donc équivalente à la formule $s_1*A*B*C$ écrite plus haut).

Enfin, lorsque la partie S' coïncide avec S (moyennage sur le facteur

S), on sera conduit à l'écriture :

$$A*B*C/S,$$

équivalente à la formule $A*B*C$ (écrite plus haut avec la "notation par défaut") ; on dira que cette nouvelle formule désigne le protocole dérivé avec la convention de la notation pleine pour le facteur S.

D'une façon générale, on parlera de notation pleine pour un facteur, lorsqu'on écrira ce facteur, en majuscule et à droite du symbole "/", pour signifier qu'on a moyenné sur ce facteur ; la notation pleine est évidemment moins concise que la notation par défaut, mais cette propriété pourra devenir un avantage (notamment dans un contexte d'emploi élargi des formules : que nous envisagerons plus loin dans ce chapitre.

L'existence de deux notations équivalentes, par défaut et pleine, pour désigner un plan de protocole dérivé conditionnel, nous fournit un exemple d'équivalences de formules. L'introduction du symbole "/" suggèrera d'autres équivalences. Reprenons, par exemple, le protocole individuel fondamental relatif au sujet s_1 , dont nous avons écrit le plan $R\langle s_1 * A * B * C \rangle$; dans cette écriture, le facteur $\{s_1\}$ apparaît comme emboîtant du facteur R ; mais $\{s_1\}$, étant un facteur constant, se trouve également croisé avec le facteur R (cf. chapitre 3), d'où l'écriture équivalente $R\langle A * B * C \rangle * \{s_1\}$, qu'on pourra aussi, en utilisant le symbole de conditionnement, écrire : $R\langle A * B * C \rangle / s_1$.

Remarque : Rappelons que nous avons défini la notion de protocole conditionnel (et illustré l'usage du symbole de conditionnement "/") en nous plaçant dans le cadre d'un protocole admettant pour plan le croisement de 2 facteurs (éventuellement composés) F et G, avec un conditionnement portant sur une partie de l'un des deux facteurs, par exemple : G/F' .

Ce cadre pourrait être indubitablement élargi, comme le suggèrera sans doute au lecteur une certaine "logique intuitive" de l'emploi du symbole de conditionnement "/". Cependant, d'une part les limites d'un tel élargissement n'apparaîtront pas si claires à cerner ; d'autre part, le cadre précédent correspond de près aux écritures admissibles dans les programmes existants, en particulier VAR3. C'est pourquoi, dans le texte présent, nous nous en tiendrons à ce cadre ; en particulier, dans le cas d'un emboîtement $F\langle G \rangle$, si G' est une partie de G (non réduite à une modalité), nous n'utiliserons pas l'écriture F/G' ; cf. à ce propos, le commentaire de la "brochure verte" (p. 33) sur le caractère superflu du symbole "/" dans les formules dérivées d'un emboîtement.

FORMULES DE PROTOCOLES DERIVES ET ECRITURES DE DERIVATION

Si nous revenons à la phase d'exploration des "données H & B": parmi les protocoles dérivés, combinant restrictions et moyennages, ceux dont le support est en correspondance bijective avec le croisement des deux facteurs expérimentaux $A_2 * B_2$, présenteront, pour l'interprétation de l'expérience, un intérêt tout particulier ; ces protocoles pourront être figurés par des graphiques, commodes pour les examens à vue : sur la figure 1, on a représenté, à titre d'exemples, les protocoles suivants, qui sont tous les trois des protocoles dérivés du protocole fondamental de groupe $S * A * B * C_2$, et qu'on pourra, à ce titre, désigner par des formules dérivées de $S * A * B * C$:

1a) restriction au sujet s_1 et moyennage sur les sessions C (ici C_2) ; d'où les formules équivalentes $A * B / s_1 * C$ (notation pleine pour C) et $A * B / s_1$ (notation par défaut pour C) ;

1b) restriction à la session c_1 et moyennage sur les sujets S ; d'où les formules équivalentes : $A * B / c_1 * S$ et $A * B / c_1$;

1c) moyennage à la fois sur les sessions (C_2) et sur les sujets S ; d'où les formules équivalentes : $A * B / C * S$, $A * B / S$, et $A * B / C$.

Remarque : la possibilité de choix entre la notation par défaut et la notation pleine (de même que celle du choix entre l'écriture indicisée ou non-indicisée pour chacun des facteurs) pourra être mise à profit pour suggérer des options méthodologiques. Ainsi, parmi les formules désignant le protocole (1c), la formule $A * B / C$ pourra être utilisée si l'on souhaite suggérer que le moyennage sur S "va de soi" mais non celui sur C (ce qui traduira), l'opposition entre facteur systématique et facteur de groupe, que nous verrons plus loin dans ce chapitre.

Ecritures de dérivation : toutes les formules écrites jusqu'à présent désignent des plans de protocoles dérivés dont elles constituent des descriptions ; mais elles peuvent être également regardées comme des écritures de dérivation, en ce sens que ces écritures conduisent directement à des procédures de calcul, comme on peut le vérifier aisément en procédant effectivement aux calculs à partir du "tableau des données" de la sortie n° 1 de l'Annexe.

Ainsi, pour obtenir le protocole $A_2 * B_2 / s_1$ (fig. 1a), on se restreint au sujet s_1 (1ère ligne du tableau) et on calcule, pour chacune des 4 combinaisons $a_1b_1, a_2b_1, a_1b_2, a_2b_2$, la moyenne des 2 valeurs correspondant aux 2 ses-

sions, d'où :

$$\begin{cases} x^{11} = 1/2 (358 + 326) & \text{(où l'indice 11 représente a1b1,} \\ x^{21} = 1/2 (389 + 348) & \text{l'indice 21 représente a2b1, etc.) (*)} \end{cases}$$

soit $\begin{cases} x^{11} = 342 \\ x^{21} = 368 \text{ etc.} \end{cases}$

[Note générale sur la présentation des résultats numériques de ce texte :
chacune des valeurs numériques est présentée après un arrondi raisonnable,
mais les calculs ont toujours été effectués sur des valeurs plus précises].

Cependant, cette possibilité d'interpréter une formule de description d'un protocole dérivé comme une écriture de dérivation n'existe ici que parce que nous nous sommes placés dans un cadre restrictif, dont nous rappelons les caractéristiques essentielles :

1) le plan à partir duquel on effectue les dérivations est un plan spécifié à l'avance, et ce plan est quasi-complet ;

2) la seule dérivation envisagée est le moyennage (ou la restriction qui peut en être regardée comme un cas particulier).

[Rappelons également les contraintes relatives aux protocoles conditionnels, sur lesquelles nous ne reviendrons pas].

Formules avec mentions : si l'on veut élargir la classe de dérivations admissibles à partir d'un plan donné (c'est-à-dire élargir le point 2) du cadre précédent), la formule de description du protocole dérivé pourra ne plus correspondre à un mode de dérivation unique. (C'est en fait déjà ce qui peut se produire avec la dérivation par moyennage, comme nous le développerons au paragraphe suivant). Si l'on veut alors des écritures qui puissent conduire à des procédures de calcul, une solution consistera à accompagner la formule de description de mentions qui préciseront le mode de dérivation.

Par exemple, pour définir le "protocole de groupe fondamental" à partir du protocole d'ensemble $R\langle S*A*B*C \rangle$, nous avons "moyenné" sur $R\langle \rangle$; mais on aurait pu préférer "prendre les médianes" sur $R\langle \rangle$; le protocole dérivé cor-

(*) Pour l'écriture des indices en position supérieure, v. le commentaire ci-dessous, l'alinéa "Protocole pondéré".

respondant serait encore décrit par la formule $S*A*B*C$; pour faire apparaître la procédure de dérivation, on pourrait alors faire suivre cette formule d'une mention telle que (MED) (pour "médiane") ; dans ce cadre élargi, la dérivation par moyennage devrait alors elle aussi être précisée par une mention telle que (MOY) (pour "moyenne") ; d'où les écritures :

$$\begin{array}{ll} S*A*B*C & \text{(MED)} \\ S*A*B*C & \text{(MOY) etc.} \end{array}$$

De telles écritures sont exploitées dans les programmes rédigés par V. Duquenne (cf. en particulier Duquenne, 1976).

Dans le texte présent, où nous privilégions les dérivations de type linéaire (cf. chapitre 6), la procédure de dérivation par "moyennage" sera toujours implicite et les seules mentions éventuelles dont nous accompagnerons les formules porteront sur le mode (équipondéré, ou pondéré) de moyennage.

Moyennage équipondéré et moyennage pondéré

En fait, même à l'intérieur du cadre restrictif adopté ici, une précision peut être indispensable pour caractériser sans ambiguïté la dérivation par moyennage : celle de la pondération à adopter.

Ainsi, pour les "données H & B" : cherchons à définir, à partir du protocole $A_2*B_2/C*s1$ (fig. 1a), le protocole (dérivé par moyennage sur A) : $B_2/A*C*s1$: ce protocole comportera deux valeurs observées x^1 et x^2 correspondant aux modalités respectives $b1$ et $b2$. On peut bien sûr prendre pour définition de x^1 la demi-somme \bar{x}^1 , c'est-à-dire :

$$\begin{array}{l} \bar{x}^1 = 1/2 (342 + 368) \text{ soit } \bar{x}^1 = 355 \\ \text{(et de la même façon } \bar{x}^2 = 368) \end{array}$$

Cependant, on peut remarquer que les deux valeurs 342 et 368 sont elles-mêmes des moyennes dont les valeurs ont été obtenues à partir d'effectif inégaux (respectivement : 71 et 24 ; cf. tableau I) et qu'en conséquence on pourrait également, pour définir x^1 , tenir compte de ces effectifs, en prenant la moyenne pondérée \hat{x}_1 définie par :

$$\hat{x}_1 = \frac{71 \times 342 + 24 \times 368}{71 + 24} \text{ , d'où } \hat{x}^1 = 349$$

et de même, en utilisant les effectifs reproduits dans le tableau I : $\hat{x}^2 = 379$.

Cette deuxième procédure sera appelée moyennage pondéré (sous-entendu : par les effectifs du protocole sur lequel on procède à la dérivation), la première procédure étant alors, par opposition, appelée moyennage équipondéré.

Pour distinguer, au niveau des écritures, les deux modes de dérivation, nous ferons suivre la formule de description du protocole dérivé de l'une des mentions (E) (pour "Equipondéré") ou (P) (pour "Pondéré") d'où les écritures :

$$\text{et } \begin{cases} B_2^c/A * C * s_1 & (E) \\ B_2^c/A * C * s_1 & (P) \end{cases}$$

Bien entendu, moyennage équipondéré et moyennage pondéré coïncident lorsque les valeurs sur lesquelles on moyenne correspondent à des effectifs égaux ; sinon, la divergence sera d'autant plus accusée que les effectifs sont plus inégaux et que les valeurs sur lesquelles on moyenne sont plus dispersées. Ainsi, dans le cas présent, la nette divergence est due d'une part aux grandes inégalités entre effectifs et d'autre part, à ce que, pour chacune des modalités de B, les valeurs sont nettement différentes.

Dans les situations où l'inégalité des effectifs rendrait très différents les résultats obtenus par l'une et l'autre procédures, les considérations de choix (que nous ne ferons qu'esquisser ici), porteront notamment :

- sur l'origine de l'inégalité des effectifs ; ici l'inégalité provient de la définition-même du facteur expérimental A qui est la "fréquence de stimulus", d'où on pourrait le cas échéant tirer argument en faveur du moyennage pondéré ;

- sur les objectifs de la dérivation par moyennage ; ici, il s'agit de définir, pour chacune des périodes préparatoires b1 et b2 une valeur qui servira essentiellement à évaluer, pour le facteur B^c, un effet "global", c'est-à-dire moyenné sur les stimulus fréquents (a1) et les stimulus rares (a2) ; or, on peut très certainement soutenir qu'au niveau de la définition de cet effet global il n'y a aucune raison de privilégier les stimulus fréquents, d'où un argument en faveur du moyennage équipondéré.

Dans la situation présente, chacune des deux procédures apparaissant défendable, il conviendra de les envisager l'une et l'autre dans les analyses, afin d'examiner si les conclusions essentielles sont sérieusement affectées (ou non) par les termes d'un choix qui ne s'impose pas. Nous y reviendrons dans les analyses des "données H & B". A propos de ces données, le

moyennage sur A sera le seul qui soulève, a priori, des problèmes de pondération - car les autres facteurs induisent des pondérations uniformes ou sensiblement uniformes.

Protocole pondéré : les considérations précédentes conduisent à généraliser la notion élémentaire de protocole (utilisé jusqu'ici) en celle de protocole pondéré, lequel par définition, sera caractérisé par la donnée d'une part, d'une famille de valeurs observées, d'autre part d'une pondération. (Les coefficients de celle-ci seront généralement, mais non obligatoirement, des effectifs, qui reflèteront le fait que le protocole considéré est dérivé d'un protocole plus fondamental). Les valeurs observées d'un protocole dérivé seront notées avec des indices supérieurs, et les coefficients de la pondération avec des indices inférieurs (l'origine de cette notation est la dualité entre "variables" et "mesures" qui jouera un rôle fondamental dans la formalisation linéaire ; cf. chapitre 6 et Réf. 1976b). Ainsi, pour un protocole de support $A_2 * B_2$, si l'on désigne par v_{11} , v_{21} , etc. les coefficients de la pondération sur $A_2 * B_2$ et par x^{11} , x^{21} les valeurs observées, on aura comme valeurs x^1 et x^2 du protocole dérivé :

$$\text{- soit } \bar{x}^1 = 1/2 (x^{11} + x^{21}) \text{ et } \bar{x}^2 = \frac{1}{2} (x^{12} + x^{22}) \text{ (moyennage équipondéré)}$$

$$\text{- soit } \hat{x}^1 = \frac{v_{11}x^{11} + v_{21}x^{21}}{v_{11} + v_{21}} \text{ et } \hat{x}^2 = \frac{v_{12}x^{12} + v_{22}x^{22}}{v_{12} + v_{22}} \text{ (moyennage pondéré)}$$

Nouvelles dérivations ; formules de protocoles dérivés d'un plan quasi-complet ("langage des plans quasi-complets").

Reprenons le problème de l'élargissement des dérivations admissibles à partir d'un plan donné (toujours supposé quasi-complet). Pour faire apparaître dans l'écriture la procédure de dérivation, on pourra (au lieu de juxtaposer, comme ci-dessus, une "mention" à la formule), envisager d'intégrer dans l'écriture de la formule elle-même l'indication de la procédure, en introduisant de nouveaux symboles. Cette solution sera particulièrement bien adaptée à la classe des dérivations que nous appelons dérivations linéaires (cf. chapitre 6), effectuées sur plan quasi-complet ; l'ensemble des formules admissibles désignant des protocoles dérivés constituera alors ce que nous appelons un langage de formules de plan quasi-complet, ou brièvement

un langage de plan quasi-complet.

Un langage de plan quasi-complet comportera :

- au minimum les 2 symboles fondamentaux "<>" et "*" (les formules écrites avec ces deux seuls symboles constituent ce que nous avons appelé les sous-formules du plan), et le symbole de conditionnement "/" ;
- de nouveaux symboles correspondant à de nouvelles dérivations ; parmi ces dernières, les plus importantes seront la dérivation intra, qui sera désignée par le symbole des parenthèses "()", et la dérivation d'interaction, désignée par le symbole du point ".".

Une définition générale de ces deux nouvelles dérivations pourra être donnée dans le cadre de la formalisation linéaire ; ici nous nous bornerons à introduire les écritures générales (proches de celles utilisées dans les programmes-machine), mais au niveau de la mise en oeuvre effective, nous nous limiterons au cas particulier (correspondant à la situation expérimentale qui sert d'illustration à notre exposé), de deux facteurs croisés, chacun à deux niveaux. En fait, dans ce cas très particulier, nous n'aurons même pas besoin de définir précisément les procédures de dérivation intra ou d'interaction, dans la mesure où nous nous intéresserons primordialement, non pas à des protocoles dérivés eux-mêmes, mais à des statistiques définies à partir de ces protocoles dérivés, lesquelles viseront à traduire les notions respectives d'effet intra et d'effet d'interaction.

Effet défini à partir d'un protocole dérivé : généralités, et effet d'un facteur à deux modalités.

Nous donnerons maintenant quelques indications sur la notion (évoquée au chapitre 1er de ce texte), d'effets liés à la structure d'un plan. Pour fixer les idées, supposons qu'à partir du protocole d'ensemble des "données H & B", nous nous posions des questions telles que les suivantes :

- "Examiner l'effet du facteur B (période préparatoire) ;
- "Examiner l'effet d'interaction entre ces facteurs expérimentaux A et B".

Nous supposerons d'abord, pour fixer les idées, qu'on envisage de traduire cette notion d'effet au moyen d'une statistique numérique.

Pour donner un sens opérationnel à de telles questions, il sera indispensable de fournir les deux précisions suivantes :

- d'une part, quel est le protocole dérivé sur lequel on choisit de définir la statistique : ce que nous appellerons le niveau d'analyse de l'effet ;

- d'autre part, quelle est la statistique que l'on choisit pour évaluer l'effet : ce que nous appellerons le critère d'évaluation de l'effet.

Dans la phase d'exploration des données, que nous envisageons d'abord, les critères d'évaluation envisagés seront des critères descriptifs (par opposition aux critères inférentiels, qui seront envisagés ultérieurement).

Parmi les effets liés à la structure du plan, on aura tout d'abord la notion d'effet d'un facteur (ou d'un facteur-partiel).

Exemple : les protocoles dérivés $B/A * C * s_1$ (fig. 1a) et $B/a_1 * C * s_1$ (cf. encore fig. 1a), permettront d'envisager respectivement l'effet du facteur B moyenné sur les facteurs A et C et restreint au sujet s_1 , et l'effet de ce même facteur restreint aux modalités a_1 et s_1 et moyenné sur les sessions C, etc. Les niveaux d'analyse respectifs de ces effets sont précisés (compte tenu de l'éventuel usage de la convention par défaut) par les termes de la formule à droite du symbole "/" (donc respectivement $A * C * s_1$ et $a_1 * C * s_1$) : On dira en outre que les effets de B précédents se situent vis-à-vis des facteurs S et C au même niveau d'analyse (restriction à s_1 et moyennage sur C) mais que vis-à-vis du facteur A, le premier est un effet global, ou moyen, (*) alors que le deuxième est un effet partiel, ou conditionnel ("pour la modalité a_1 ").

Quant au critère d'évaluation : dans le cas particulier d'un facteur à deux modalités (comme ici le facteur B_2), il sera naturel en ordonnant conventionnellement les 2 modalités, de prendre comme statistique la différence (orientée) $x^1 - x^2$ (x^1 et x^2 désignant les valeurs observées du protocole) ; nous appellerons cette statistique l'effet observé du facteur à deux niveaux.

Du point de vue des écritures, on pourra faire figurer, à gauche de la formule du protocole, la spécification de la statistique à calculer : ici l'effet observé. Ainsi, on écrirait (abrégeant "effet observé en EFF.OBS) :

(*) Le qualificatif de "moyen" sera tout à fait adéquat dans toutes les situations où, comme ici, l'effet global apparaît comme une moyenne d'effets conditionnels.

Exemples numériques: l'effet de B_2 pour le niveau d'analyse $a_1 * C_2 * s_1$ (cf. fig. 1a) a pour valeur : $342 - 372 = -30$. De même, l'effet de B_2 pour le niveau d'analyse $a_2 * C_2 * s_1$ a pour valeur $368 - 401 = -32$ (après arrondi). Quant à l'effet de B_2 pour le niveau d'analyse $A * C_2 * s_1$, il met en jeu un moyennage sur A, donc, a priori, on devra préciser le mode de moyennage utilisé ; les résultats sont les suivants :

- selon le moyennage équipondéré sur A : $355 - 386$, soit -31 ;
- selon le moyennage pondéré sur A : $349 - 379$, soit encore -31 (valeurs arrondies).

[Dans le cas présent, le mode de dérivation, qui affectait les valeurs elles-mêmes du protocole $B_2 / A * C * s_1$, n'affecte donc pratiquement pas l'effet calculé à partir de ces valeurs (*) ; ce qui est une constatation encourageante !].

Croisement de deux facteurs à 2 niveaux : effets intra et effet d'interaction

Nous introduisons maintenant, dans le cas particulier du croisement de 2 facteurs à 2 niveaux : $A_2 * B_2$, les notions d'effet intra et d'effet d'interaction.

1) Effet intra

En fait la notion d'effet (intra) d'un facteur (disons B) à l'intérieur d'une modalité (disons a_1) de l'autre facteur coïncide avec la notion d'effet conditionnel de ce facteur (B) pour cette modalité (a_1) de l'autre facteur. (Les deux notions ne se distingueront que lorsqu'on définira les effets intra et conditionnel relatifs à une partie comportant plusieurs modalités : cf. chapitre 6). Ainsi, l'effet de B à l'intérieur de la modalité a_1 pour le niveau d'analyse $C * s_1$, peut être écrit, à l'aide du symbole des parenthèses "()" EFF.OBS. $B(a_1) / C * s_1$. Cet effet coïncide avec l'effet de B restreint à la modalité a_1 , et également moyenné sur C et restreint à s_1 , qui s'écrit :

$$\text{EFF.OBS } B/a_1 * C * s_1$$

2) Effet d'interaction

Reprenons le croisement $A_2 * B_2$. Pour l'un des facteurs (disons B), on

(*) Comme on pourra le vérifier, ce "phénomène" se produira, dans un plan $A_2 * B_2$, notamment chaque fois que, comme ici, le plan est orthogonal (effets proportionnels) et que les deux effets conditionnels sont voisins (c'est-à-dire l'effet d'interaction proche du zéro).

peut définir un effet de ce facteur à l'intérieur de chacune des modalités de l'autre facteur, donc ici les effets intra B(a1) et B(a2). Lorsque chacun des 2 facteurs est à 2 niveaux, on définira l'effet d'interaction comme la différence (orientée) de ces deux effets intra. Par exemple, l'effet d'interaction entre A₂ et B₂, pour le niveau d'analyse C*s1, sera défini comme la différence des deux effets intra B(a1)/C*s1 et B(a2)/C*s1. La notation générale pour désigner l'interaction étant le symbole point ".", nous écrirons donc ici :

$$\text{EFF.OBS } A_2.B_2/C*s1$$

(numériquement : cet effet a ici pour valeur : -30 - (-32) = +3 (après arrondi)).

Avec les notations introduites plus haut pour un protocole A₂*B₂, l'effet d'interaction entre 2 facteurs, chacun à 2 niveaux, sera donc défini comme (x¹¹ - x¹²) - (x²¹ - x²²) = (x¹¹ - x²¹) - (x¹² - x²²) ; la notion d'interaction est symétrique, et les deux écritures A.B et B.A seront donc synonymes (propriété qui restera vraie pour la notion générale d'interaction).

Sur les graphiques, l'effet d'interaction correspond à l'écart au parallélisme des deux segments relatifs à b1 et à b2 (donc + 3 pour le protocole A₂*B₂/C*s1). On trouverait de même, pour le protocole A₂*B₂/C₂*S (figure 1c), un effet d'interaction égal à -7, etc.

Remarque : A partir du facteur A₂*B₂ = {a1b1, a2b1, a2b2}, on peut définir, par regroupement des modalités a1b1 et a2b2, d'une part, et des modalités a2b1 et a1b2, d'autre part, le nouveau facteur à deux niveaux {a1b1 a2b2, a2b1 a1b2} ; et l'effet de ce nouveau facteur pourra être défini :

$$\left\{ \begin{array}{l} \text{soit comme } \frac{x^{11} + x^{22}}{2} - \frac{x^{21} + x^{12}}{2} \quad (\text{définition équilibrée}) ; \\ \text{soit comme } \frac{v_{11}x^{11} + v_{22}x^{22}}{v_{11} + v_{22}} - \frac{v_{21}x^{21} + v_{12}x^{12}}{v_{21} + v_{12}} \quad (\text{définition pondérée}). \end{array} \right.$$

Ainsi, l'effet d'interaction entre A₂ et B₂ coïncide avec le double de l'effet de ce nouveau facteur défini selon la définition équilibrée.

VERS LA NOTION D'ANALYSE LOCALE

Les effets observés que nous avons définis (en nous plaçant dans des situations très simples) fournissent des exemples de procédures applicables à des protocoles dérivés d'un plan quasi-complet. Lorsque on envisagera des critères d'évaluation plus complexes (*), l'opérationnalisation de la notion d'effet pourra conduire à des procédures elles aussi plus complexes, allant souvent au-delà du calcul d'une seule statistique numérique. Nous dirons, dans ce contexte élargi, que la mise en oeuvre d'une procédure (sur le protocole dérivé) constitue une analyse locale (du protocole primitif). La donnée d'une formule de protocole dérivé et de la spécification de la procédure à appliquer sera appelée une demande d'analyse locale.

Le terme d'analyse "locale" ne devra donc pas prêter à malentendu, en suggérant une analyse qui porterait nécessairement sur une partie stricte du protocole, mais plutôt évoquer l'idée que les données sont condensées "d'un point de vue particulier" (par exemple, celui de l'interaction entre A et B, à un niveau d'analyse spécifié).

Les notions d'effet et l'analyse locale, que nous introduisons ici d'une façon très intuitive, pourront être formalisées :

- tout d'abord, au niveau des structures linéaires (cf. chapitre 6 et Réf. 1976b) : l'examen des données "selon un point de vue particulier" correspondra à une projection du vecteur des observations sur un sous-espace (la notion d'analyse générale de son côté, correspondant à celle de décomposition d'un espace vectoriel en sous-espaces orthogonaux) ;

- ensuite au niveau de la statistique bayésienne (cf. Réf. 1977d)

Analyses locales et analyse générale liée à un facteur composé.

La notion qu'il conviendra d'opposer à celle d'analyse locale sera celle d'analyse générale liée à un facteur composé ; une analyse générale est constituée de plusieurs analyses locales visant à examiner les données selon plusieurs points de vue, à la fois "complémentaires" et "aussi peu redondants" que possible. Comme exemples, nous donnerons les décompositions des effets liés au croisement de 2 facteurs $A*B$. (On pourra se placer dans le cas de facteurs à seulement 2 modalités, mais les notions et notations qui suivent

(*) V. par exemple (quitte à anticiper quelque peu) les résultats présentés dans le tableau III).

se généralisent à des facteurs croisés à un nombre quelconque de modalités).

Le croisement $A \times B$ permet de définir des effets de divers types : effets globaux (ou moyens), effets conditionnels, effet d'interaction ; une analyse générale du croisement $A \times B$ sera définie par l'une des décompositions suivantes :

- | | | |
|--|---|---------------------------------|
| 1) effet global de A : A/B | } | (décomposition symétrique) |
| effet global de B : B/A | | |
| effet d'interaction : A.B | | |
| 2) effets conditionnels A/b, ou intra A(b) | } | (décompositions dissymétriques) |
| (b parcourant B) et l'effet global B/A | | |
| 2') (en permutant les rôles de A et B) | | |

La décomposition 1) sera appelée décomposition canonique des effets liés au croisement $A \times B$. Méthodologiquement, cette décomposition sera privilégiée si (et seulement si) les deux facteurs A et B jouent vis-à-vis des objectifs de l'expérience des rôles symétriques ; mais, lorsque vis-à-vis de ces objectifs, les facteurs sont hiérarchisés, le facteur A jouant le rôle de facteur principal, et le facteur B étant subordonné, c'est la décomposition 2) qui sera préférée aux autres.

Comme nous l'avons dit, la notion d'effet conditionnel A/b coïncide avec celle d'effet intra-modalité A(b). Quant à la notion générale d'effet A intra B, on pourra, pour l'introduire intuitivement, considérer la famille des effets intra-b, c'est-à-dire ici :

$$\begin{cases} A(b_1) \\ A(b_2) \end{cases}$$

On voit alors que les décompositions 1) et 2) peuvent être obtenues à partir de la première décomposition suivante des effets liés à $A \times B$:

$$\begin{cases} A(B) : \text{effet de A intra-B} \\ B/A : \text{effet global de B,} \end{cases}$$

en redécomposant A(B) de deux manières :

- 1°) $\begin{cases} A/B \\ A.B \end{cases}$, d'où la décomposition 1) ;
- 2°) $\begin{cases} A/b_1 \\ A/b_2 \end{cases}$, d'où la décomposition 2).

Décomposition canonique des effets liés à un croisement, à un facteur quasi-complet

La décomposition 1) ci-dessus s'étendra immédiatement au croisement de plusieurs facteurs ; par exemple, on aura, pour le croisement $A \times B \times C$ de 3 facteurs :

- les 3 effets globaux : $\left\{ \begin{array}{l} A \text{ (ou } A/B \text{ ou } A/C, \text{ ou } A/B+C) \\ B \text{ (ou } B/A \text{ etc.)} \\ C \text{ (ou } C/A \text{ etc.)} \end{array} \right.$

- les 3 effets d'interaction "binaires" (entre facteurs 2 à 2) appelés encore effets d'interaction "simples" (car un seul symbole ".") :

$\left\{ \begin{array}{l} A.B \text{ (ou } A.B/C) \\ A.C \text{ (ou } A.C/B) \\ B.C \text{ (ou } B.C/A) \end{array} \right.$

- l'interaction "ternaire" (entre 3 facteurs) appelés encore "double" (car 2 symboles ".") : $A.B.C$.

Une telle décomposition sera appelée décomposition canonique des effets liés au croisement. Plus généralement, à tout facteur quasi-complet on peut associer une décomposition canonique des effets liés à ce facteur. Cette décomposition des effets, bien entendu, est étroitement apparentée à la décomposition canonique de la formule de facteur quasi-complet mentionnée au chapitre III ; (un algorithme, dû à V. Duquenne, permet d'obtenir l'une et l'autre) ; lorsque le contexte ne permettra pas d'ambiguïté, on pourra parler simplement de décomposition canonique du facteur quasi-complet.

{ [Dans les programmes VAR3 et VAR4, la décomposition canonique d'un sous-facteur du plan déclaré pourra être engendrée par une demande unique d'analyse générale ; on trouvera dans la sortie n° 2 de l'Annexe un exemple de décomposition canonique du croisement de 3 facteurs. Cependant, d'un point de vue méthodologique, la décomposition canonique sera loin d'être toujours privilégiée].

A titre d'illustration, nous reprendrons les exemples de facteurs quasi-complets dont nous avons donné (à la fin du chapitre III) la décomposition canonique en sous-formules ; et nous donnerons maintenant pour chacun d'eux, la décomposition canonique des effets.

Formule	Décomposition canonique des effets
$E\langle F \rangle$	$F ; E(F).$
$R\langle E\langle F \rangle \rangle$	$F ; E(F) ; R(E\langle F \rangle).$
$A * B$	$A ; B ; A.B .$
$E\langle F \rangle * D$	$F ; D ; F.D ; E(F) ; E(F).D .$
$R\langle E\langle F \rangle * D \rangle$	$F ; D ; F.D ; E(F) ; E(F).D ; R(E\langle F \rangle * D).$
$R\langle E\langle F \rangle * D \rangle * A$	$F ; D ; F.D ; E(F) ; E(F).D ; R(E\langle F \rangle * D) ;$ $A ; F.A ; D.A ; F.D.A ; E(F).A, E(F).D.A ;$ $R(E\langle F \rangle * D).A$

N.B. (1) On remarquera les rôles distincts que jouent, dans les formules précédentes, les chevrons " $\langle \rangle$ " et les parenthèses $()$.

(2) Au lieu du symbole des parenthèses on aurait pu, pour désigner l'effet intra, choisir un autre symbole (tel que " $\subset \supset$ "), le choix du symbole des parenthèses est possible ici précisément parce que l'écriture des formules ne nécessite aucun "vrai" symbole parenthétique, dont la seule valeur serait celle de séparateur (lorsque, dans ce texte, nous avons eu besoin, pour des raisons essentiellement didactiques, d'un tel symbole, nous avons utilisé les crochets $[]$).

NOUVEL ELARGISSEMENT DE L'USAGE DES FORMULES : EXPRIMER DES "QUESTIONS"

A POSER AUX DONNEES

Jusqu'à présent, nous en sommes toujours restés au cadre restrictif (précisé par le point 1), p. 43, d'un plan quasi-complet posé au départ, avec des formules qui désignent des protocoles dérivés de ce plan. Mais par delà (*) ce cadre restrictif, on pourrait songer à conférer une plus large autonomie à l'emploi des formules, notamment à utiliser une formule pour exprimer une "question" que l'on souhaite poser aux données, sans chercher nécessairement à spécifier complètement ni le protocole dérivé, ni même le protocole primitif. Ainsi, envisageons les questions suivantes, formulées de façon verbale;

- "examiner l'effet du facteur B moyenné sur l'ensemble des sessions C"
- "examiner l'effet du facteur B restreint à la session c1".

Ces questions pourraient être exprimées par les formules respectives : B/C et B/c1 ; ces formules n'étant considérées, au moins dans un premier temps, que comme de simples "traductions" des formulations verbales.

De même, la question :

- "examiner l'interaction entre les facteurs A et B, moyennée sur l'ensemble des sessions C"

pourrait être exprimée par la formule : A.B/C, etc.

Mais dans un deuxième temps, pour tout protocole décrit par un plan (quasi-complet) tel que la formule de la question soit une formule en langage engendré par ce plan (exemples de tels plans : $S * A * B * C$, $R < A * B * C > / s1$), la formule pourrait caractériser un protocole dérivé de ce protocole.

Dans ce cadre élargi, il deviendra utile de distinguer les cas où une formule exprimant une question est relative à un protocole spécifié ; nous parlerons alors de formule d'interrogation. A titre d'illustration, on a représenté dans le tableau III deux questions (lignes du tableau), lesquelles appliquées d'une part au protocole de groupe fondamental, d'autre part au protocole individuel fondamental du sujet n° 1 (colonnes du tableau), conduisent aux formules d'interrogation écrites dans les cases du tableau.

(*) et également, au cours de la démarche, antérieurement à l'analyse des données, c'est-à-dire au moment de la planification de l'expérience, où on pourra poser des relations entre facteurs sans un plan fixé au départ.

Malgré tout l'intérêt que pourrait présenter l'élargissement notionnel que nous venons d'évoquer, nous nous en tiendrons, dans le présent texte, au cadre restrictif indiqué plus haut. Ce cadre correspond à celui des programmes-machine, dans lesquels l'utilisateur doit fournir un "plan d'analyse déclaré" par lui, les demandes d'analyse effectuées ensuite portant sur des protocoles dérivés de ce plan déclaré. Cependant, afin de suggérer la possibilité d'un cadre élargi, nous tendrons à privilégier, dans l'écriture d'une formule, la notation pleine, laquelle dans le cadre élargi, présenterait, par rapport à la notation par défaut, l'avantage de mieux carner les protocoles pour lesquels cette formule désignerait un protocole dérivé : par exemple, l'écriture A.B/C (avec la notation pleine pour C) ne saurait être applicable qu'à des protocoles qui admettent C comme facteur (tels que $R<s1*A*B*C>$, ou $S*A*B*C$, etc.), et exclura donc des protocoles ayant déjà fait l'objet d'une restriction sur C (tels que $S*A*B/c1$, etc.) ; alors que l'interprétation d'une notation par défaut (telle que A.B) exige manifestement que soit spécifié le protocole à partir duquel on dérive.

DE L'EXPLORATION DES DONNEES AUX ANALYSES FINES

Si l'on procède à l'exploration des protocoles individuels des données "H & B", on constate que les effets individuels sont tantôt positifs, tantôt négatifs, leurs valeurs absolues s'échelonnant entre 3 (sujet s1) et 35 (sujet s12). L'effet moyen très faible (- 7 ms) calculé à partir du protocole de groupe (figure 1c) résulte donc d'une "compensation". En fait, chacun des effets individuels est déjà lui-même un effet moyen (obtenu par dérivation sur le facteur Répétitions).

[Numériquement : pour le sujet n° 1, l'écart-type-corrigé (*) des effets d'interaction, à l'intérieur des 8 conditions, est de l'ordre de 65, la moyenne des 12 écarts-types-corrigés intra-individuels étant de l'ordre de 75].

A la variabilité intra-individuelle se superpose, pour les effets calculés à partir du protocole de groupe, la variabilité interindividuelle. Numériquement : l'écart-type corrigé des effets d'interaction individuels est

(*) Pour une famille $(x_i, i \in I)$ de n observations numériques, de moyenne $m = \frac{\sum x_i}{n}$, nous appelons variance-corrigée la statistique $s^2 = \frac{\sum (x_i - m)^2}{n-1}$ ("corrignée", par rapport à la variance "vraie" $\frac{\sum (x_i - m)^2}{n}$, pour tenir compte des degrés de liberté : $n-1$ au lieu de n) et l'écart-type corrigé sa racine-carrée s .

ici de l'ordre de 20.

Une telle situation peut être considérée comme tout à fait représentative de nombreuses recherches expérimentales actuelles (du moins dans le domaine de la psychologie) : l'expérimentaliste se pose des "questions fines" qui l'amènent à recourir à des "analyses fines", lesquelles iront plus loin que les examens à vue d'effets moyens, en faisant entrer en compte les diverses variabilités. Parmi les "questions fines" les plus courantes, on mentionnera : l'examen d'un schéma additif (ou ce qui revient au même, l'examen d'une interaction) comme dans l'expérience analysée ici ; l'examen d'un schéma de régression, etc.

Sans que l'opposition avec la phase d'exploration des données ait rien d'absolu, nous dirons que dans la phase des "analyses fines" :

- les interrogations seront en nombre limité, directement liées aux objectifs de la recherche ;

- les critères d'évaluation renverront à un cadre susceptible de conduire à des conclusions précises de caractère inductif, au moyen de procédures inférentielles : ce cadre comportera au moins un modèle d'échantillonnage. Quant au choix des niveaux d'analyse, il deviendra tout à fait crucial, au moins autant que celui des critères d'évaluation. En règle générale, une interrogation à un niveau d'analyse "grossier" ne permettra que de cerner d'assez loin la "véritable interrogation" que l'on souhaite poser aux données (par exemple, ici : les interrogations au niveau du groupe). Mais à l'opposé, des interrogations posées à un niveau trop "fin" risqueront d'entraîner une "incapacité à conclure" de la part des procédures statistiques, faute d'une précision expérimentale suffisante (par exemple : les interrogations individuelles).

Dans ce texte, la phase des "analyses fines" des données expérimentales consistera essentiellement en un ensemble planifié d'analyses locales approfondies (généralement inférentielles).

Cet accent que nous mettons ici sur les analyses locales, aussi bien lors de la phase des analyses fines que lors de la phase d'exploration, vise à satisfaire une exigence primordiale déjà évoquée à plusieurs reprises : permettre à l'expérimentaliste d'interroger ses données d'une façon aussi efficace et précise que possible, dans les seules limites du plan utilisé. Dans

le prochain chapitre, nous commenterons les implications méthodologiques de cette perspective d'"analyse locale des données".

STRUCTURE STATISTIQUE D'UN PLAN

Facteur de groupe et facteur systématique

Dans les analyses effectuées sur le protocole $S \times A \times B \times C$, le facteur S jouera un rôle particulier : ses modalités (les sujets) seront toujours traitées d'une façon symétrique (cf. Réf. 1972-73) ; nous dirons que le facteur S est un facteur de groupe, les facteurs A,B,C étant, par opposition, qualifiés de facteurs systématiques. (On s'intéresse, en principe, aux modalités spécifiques de ces facteurs).

De même, pour un protocole individuel $R \langle A \times B \times C \rangle$, les modalités du facteur R à l'intérieur de chacune des combinaisons de $A \times B \times C$ seront toujours traitées de façon symétrique ; nous dirons encore que le facteur R (ou plus explicitement : le facteur $R \langle \rangle$) est un facteur de groupe.

Un plan dans lequel chaque facteur élémentaire sera caractérisé soit comme facteur de groupe, soit comme facteur systématique, sera dit muni d'une structure statistique.

(Dans les dérivations effectuées antérieurement, nous avons déjà, implicitement, tenu compte des structures statistiques précédentes : lorsque, par exemple, nous avons, à partir du protocole $S \times A \times B \times C$, moyenné sur toutes les modalités du facteur S).

Structures statistiques remarquables : $S \langle G \rangle$ (groupes séparés), $S \times T$ (groupes appareillés) et structure $S \langle G \rangle \times T$.

Parmi les structures statistiques, on distinguera les trois structures "remarquables" notées $S \langle G \rangle$, $S \times T$ et $S \langle G \rangle \times T$.

La structure $S \langle G \rangle$ est celle des groupes séparés (ou groupes indépendants) ; elle est définie par l'emboîtement d'un facteur de groupe dans un facteur systématique. Un cas typique est celui de "sujets S emboîtés dans des groupes G", (le facteur G pouvant résulter de la description des sujets au moyen soit d'un facteur de classification, soit d'un facteur décrivant l'affectation des sujets à un ensemble de conditions expérimentales, etc.) ;

d'où la notation générique $S\langle G \rangle$, qui servira à désigner cette structure. (N.B. : dans la structure $S\langle G \rangle$, c'est bien S qui est le "facteur de groupe", alors que l'ensemble des "groupes" G est un facteur systématique). Mais bien entendu, le facteur de groupe pourra être tout autre qu'un ensemble de sujets. Par exemple, dans les données "H & B", chaque protocole fondamental individuel $R\langle A*B*C \rangle$ sera considéré comme muni de la structure statistique des groupes séparés ; le facteur de groupe emboîté sera le facteur Répétitions $R\langle \rangle$, et le facteur systématique emboîtant le croisement $A*B*C$.

La structure $S*T$ est celle des groupes appareillés ; elle est définie par le croisement d'un facteur de groupe avec un facteur systématique. Un cas typique est celui de "Sujets S croisés avec des traitements T", d'où la notation générique $S*T$ pour désigner cette structure. Par exemple, le protocole fondamental $S*A*B*C$ sera muni de la structure statistique des groupes appareillés ; le facteur de groupe croisé sera le facteur Sujets S, et le facteur systématique croisé, le croisement $A*B*C$ (facteur Conditions).

La structure $S\langle G \rangle*T$ sera définie comme la combinaison des structures $S\langle G \rangle$ et $S*T$: le facteur de groupe (S) est emboîté dans un facteur systématique (G) et croisé avec un autre facteur systématique (T). On obtiendrait un exemple de cette structure, par exemple si les sujets de l'expérience "H & B" avaient été décrits au moyen d'un nouveau facteur systématique E (par exemple : facteur de classification ou ensemble de tâches préliminaires), d'où, pour le protocole fondamental, le plan $S\langle E \rangle*A*B*C$, dont la structure statistique serait de type $S\langle G \rangle*T$.

Par contre, le plan d'ensemble des "données H & B" : $R\langle S*A*B*C \rangle$, ne sera pas muni de la structure statistique $S\langle G \rangle*T$, car nous avons considéré que les facteurs S et $R\langle \rangle$ étaient tous les deux des facteurs de groupe. La structure $S\langle G \rangle*T$ (avec un seul facteur de groupe) sera néanmoins suffisante pour les traitements que nous envisageons dans ce texte, dans la mesure où toutes les questions envisagées seront posées au niveau de protocoles dérivés relevant de la structure $S\langle G \rangle*T$: soit qu'il s'agisse de chacun des protocoles fondamentaux individuels, soit du protocole de groupe fondamental.

Remarque : les structures statistiques $S*T$ et $S\langle G \rangle*T$ constituent la formalisation des types de plans que les expérimentalistes appellent souvent "plans à mesure répétées" - qu'on peut appeler également : "plans à plusieurs observations sur la même variable".

Facteur adjoint, effet adjoint

A ce point de l'exposé, il nous faut mentionner (même si nous ne faisons que les évoquer, dans le cadre de la structure statistique $S\langle G \rangle * T$) les notions de facteur adjoint et d'effet adjoint.

Etant donné un facteur systématique d'une structure $S\langle G \rangle * T$, on appellera facteur adjoint minimal de ce facteur, le facteur composé de ce facteur avec le facteur de groupe, et facteur adjoint (en général) tout facteur admettant le facteur adjoint minimal comme sous-facteur (ces définitions s'étendent immédiatement aux facteurs-partiels).

Exemples :

- dans le plan $R\langle A * B * C \rangle$, muni de la structure statistique $S\langle G \rangle$, le facteur $A * B$ admet $R\langle A * B \rangle$ comme facteur adjoint minimal et $R\langle A * B * C \rangle$ comme facteur adjoint (maximal) ;

- dans le plan $S * A * B * C$, muni de la structure statistique $S * T$, le facteur $A * B$ admet $S * A * B$ comme facteur adjoint minimal et $S * A * B * C$ comme facteur adjoint (maximal).

La notion de facteur adjoint conduira à celle d'effet adjoint (à titre indicatif : dans la structure $S\langle G \rangle$, l'effet adjoint du facteur G sera l'effet intra $S(G)$; dans la structure $S * T$, l'effet adjoint du facteur T sera l'effet d'interaction $S.T$). Dans une analyse locale, portant sur l'examen d'un certain effet, l'effet adjoint de l'effet examiné (ou l'un des effets adjoints) pourra servir de "terme de référence" à l'effet examiné, et à ce titre intervenir dans des procédures diverses : tout d'abord, dans les analyses en termes de tests de signification, où les effets adjoints fourniront les divers dénominateurs envisageables pour les "rapports F " (cf. chapitre 5) ; mais également, dans des procédures comme les analyses fiduciaires (cf. chapitre 7).

LANGAGE D'INTERROGATION ET LANGAGE DE COMMANDE DE PROGRAMMES-MACHINE

La mise au point d'un langage utilisable comme langage d'interrogation conduit naturellement à concevoir des programmes-machine d'analyse statistique dans lesquels le langage sera utilisé comme langage de commande, c'est-à-dire dans lesquels une formule pourra servir d'entrée comme demande

d'analyse de la part de l'utilisateur, et engendrer automatiquement, de la part du programme, un certain nombre de procédures.

Dans la phase d'exploration des données, l'intérêt pratique de tels programmes apparaîtra lorsque le plan du protocole est relativement complexe et que l'expérimentaliste désire procéder à un grand nombre de dérivations, même si chacune d'elles conduit à des calculs simples.

Dans la phase des "analyses fines", l'intérêt des programmes résidera plutôt dans la complexité des analyses effectuées.

C'est dans cette perspective qu'ont été élaborés plusieurs programmes de complexité diverses, aussi bien du point de vue de la syntaxe du langage que de la nature des dérivations effectuées.

Les principaux programmes expressément conçus en vue de l'exploration des données sont ceux de V. Duquenne ; ces programmes mettent en oeuvre des versions du langage des plans quasi-complets, et engendrent les tableaux et statistiques descriptives courantes (cf. V. Duquenne, 1976).

Les programmes de la série VAR ont été conçus comme des programmes d'analyse fine (inférentielle pour VAR3) ; les derniers en date de ces programmes, VAR3 et VAR4, mettent en oeuvre le langage des comparaisons, qu'on peut regarder comme une version enrichie de ce que nous avons appelé ici un langage de plan quasi-complet (ce langage est exposé en détail dans la "brochure verte", Réf. 1976a).

Nous signalerons enfin, pour son intérêt didactique, le programme BORDAU qui met en oeuvre le langage des plans complets (comportant les deux seuls symboles "*" et "/") ; langage extrêmement élémentaire mais non trivial, d'où l'intérêt de ce programme à des fins d'initiation (la notice technique du programme BORDAU se trouve dans la Réf. 1977c).