

PROGRAMMES DE DESCRIPTION ET D'ANALYSES INFÉRENTIELLES  
DE DONNÉES EXPERIMENTALES POUR MICRO-ORDINATEURS

Le développement actuel des micro-ordinateurs est susceptible d'introduire de profondes modifications dans les conditions d'expérimentation. En autorisant, en particulier, l'expérimentation en temps réel et le recueil direct d'une masse d'informations (temps de réactions, conversion numérique de données analogiques, etc...), les micro-ordinateurs devraient permettre d'accroître considérablement les possibilités de recherche, tout en diminuant les risques d'erreurs liés à la médiation humaine dans le codage et le transfert des données.

Malheureusement, l'intérêt potentiel de cet apport est grandement réduit par les difficultés que posent l'analyse et le traitement des données recueillies. Il est le plus souvent nécessaire de réaliser des programmes spécifiques pour les analyses descriptives élémentaires et les transformations préliminaires à une analyse inférentielle, telles que moyennage ou transformation d'échelle. Quant à l'analyse inférentielle elle-même, les possibilités sont réduites par la rigidité des programmes habituellement proposés. Ainsi, les multiples programmes d'analyse de variance pour micro-ordinateur publiés dans "*Behavior Research Methods and Instrumentation*" ont pour caractéristiques communes d'être limités tant en puissance (ils admettent au maximum quatre facteurs systématiques) qu'en souplesse (aucun ne permet l'exécution d'analyses partielles ou de test de tendance) ; de plus, chaque programme correspond à un plan expérimental déterminé (par exemple : un facteur inter-sujets et deux facteurs intra-sujets), ce qui conduit à un travail de copie et de vérification assez important.

Une solution alternative consiste à transférer les données sur un ordinateur puissant permettant l'utilisation de programmes standards polyvalents. Cette solution est évidemment la meilleure lorsque le laboratoire dispose, d'une part du matériel permettant un transfert direct des données, et d'autre part d'opérateurs spécialisés. Dans le cas contraire, qui est le plus fréquent, ce genre d'opération est long, coûteux, et constitue une source potentielle d'erreurs. Une grande part de l'avantage des micro-ordinateurs, qui se trouvent aujourd'hui, de par leur coût

réduit, à la portée de petits laboratoires, se trouve ainsi perdue.

Les programmes ici proposés tentent de mettre à la disposition des utilisateurs de micro-ordinateurs plusieurs des possibilités offertes par les programmes développés sur des systèmes importants, et en particulier par les programmes de description et d'analyse inférentielle dérivés des travaux de H. Rouanet de D. Lépine sur la planification et le traitement des données expérimentales (cf. en particulier, Rouanet & Lépine, 1977) [2].

Trois programmes, utilisables séquentiellement, ont été réalisés.

Le premier (VARE) est un programme d'entrée et de stockage des données, pour le cas où celles-ci ne sont pas recueillies directement en mémoire.

Le deuxième programme (VARD) permet, d'une part, d'opérer des dérivations de protocoles (supprimer une ou plusieurs modalités d'un ou plusieurs facteurs, moyennage sur un facteur, transformation d'échelle de mesure), d'autre part, de fournir différents modes de description du protocole original ou des protocoles dérivés (établissement de distributions, calcul de moyennes par cellule, par modalité de chaque facteur, par croisement de modalités, calcul de variance, etc.).

Le troisième programme (VARI) effectue les analyses inférentielles. Outre la décomposition canonique associée au plan, ce programme autorise les conditionnements, les comparaisons partielles planifiées, et les tests de tendance de degré quelconque.

Dans le cas de comparaison à un seul degré de liberté, la distribution fiduciaire peut être immédiatement dérivée de l'effet observé (fourni par VARD) et du rapport F (fourni en VARI) (Lecoutre, 1977) [2].

Soit, pour concrétiser ces différentes possibilités, un plan de la forme :  $\underline{S} < G_3 * M_2 > * E_{10} * R_2$ . Trois souches de souris de génotypes différents ( $G_3$ ) comprenant des mâles et des femelles ( $M_2$ ), sont soumises à une séance d'apprentissage comprenant 10 blocs d'essais ( $E_{10}$ ), deux fois consécutives ( $R_2$ ). Supposons que l'on souhaite supprimer l'une des souches, mélanger mâles et femelles des deux souches restantes, et analyser comment celles-ci diffèrent dans l'apprentissage, en première séance. Après l'examen des statistiques descriptives pertinentes, un test inférentiel possible sera le test d'interaction entre la tendance linéaire évaluée sur  $E$  d'une part, et  $G$  d'autre part, après moyennage sur  $M$ , et restriction aux deux modalités de  $G$  et à la première modalité de  $R$ . Toutes ces opérations sont facilement réalisables. Les demandes s'effectuent sur un mode interactif. Le choix d'un traitement donné s'opère en répondant séquentiellement aux questions posées, qui sont formulées dans un langage explicite. L'apprentissage d'un code arbitraire n'est

donc pas requis ; seule est nécessaire la formalisation du plan en termes de croisements et d'emboitements de facteurs. Pour minimiser les risques d'erreurs provenant d'une inattention de l'opérateur dans le choix des dérivations, la formule du plan d'analyse s'inscrit automatiquement avant chaque résultat ; dans l'exemple cité, elle serait  $\underline{S} < G_2 > * E_{10}$ , les facteurs M et R n'étant plus source de variations.

Les programmes proposés présentent différentes limitations par rapport aux programmes issus des travaux de H. Rouanet et de D. Lépine, actuellement implantés sur IBM 370.

Les unes sont relatives au type de protocole analysable. Le nombre maximum de données d'une analyse dépend des capacités de mémoire du micro-ordinateur ; VARD et VARI utilisent respectivement environ 5 K et 6 K octets, ce qui permet de traiter, si l'on dispose de 48 Koctets de RAM, des fichiers de 4000 à 6000 données. Le nombre de facteurs n'est pas limité, et leur répartition en facteurs "inter-sujets" (plan indépendant) et "intra-sujet" (plan appareillé) peut être quelconque. Mais il est nécessaire d'opérer sur des plans totalement équilibrés : les observations doivent être réparties également dans les différentes cellules. Sous cette condition, n'importe quel plan quasi-complet peut être traité en VARD. L'entrée en VARI suppose une condition supplémentaire : l'absence d'emboitements entre facteurs systématiques.

D'autres limitations sont relatives aux traitements réalisables sur un protocole donné. Un certain nombre de statistiques ne peuvent être obtenues qu'après plusieurs passages en machine et/ou quelques calculs supplémentaires. C'est en particulier le cas lorsque plusieurs carrés moyens peuvent estimer une source de variation aléatoire ; de façon générale, un seul F, utilisant la source adjointe minimale, est directement calculé. De même, les comparaisons partielles à plusieurs degrés de liberté impliquant conjointement des regroupements entre modalités exigent un nombre de passages en machine égal au nombre de degrés de liberté.

Il faut signaler enfin que les traitements sont souvent longs, la rapidité d'exécution ayant été sacrifiée au profit de la souplesse et de la diversité des opérations réalisables.

Il apparaît que, pour un très large éventail de protocoles requérant l'analyse des comparaisons, beaucoup des limitations décrites ne sont pas réellement pénalisantes. La propriété la plus restrictive, sur laquelle il convient d'insister, est sans doute l'existence d'un plan expérimental équilibré.

Ces programmes sont écrits en BASIC pour APPLE II. La transposition à d'autres systèmes implique que la dimension des vecteurs ne soit pas limitée (comme dans le WANG 2200 par exemple), toutes les données étant stockées sur un vecteur unique. Les listings, et une notice d'utilisation (11 p.) peuvent être obtenus auprès de l'auteur ; les utilisateurs d'APPLE II ou d'ITT 2020 peuvent envoyer une disquette vierge pour copie directe (spécifier DOS 3.2 ou 3.3).

#### BIBLIOGRAPHIE

- [1] LECOUTRE, B., Note sur le calcul de la distribution fiduciaire pour une inférence sur un contraste entre moyennes. *Cahiers de Psychologie*, 1977.
- [2] ROUANET, H., LEPINE, D., Introduction à l'analyse des comparaisons pour le traitement des données expérimentales. *Informatique et Sciences Humaines*, 1977, n° 33-34, p.11-125.

P. PERRUCHET

---

Laboratoire de Psychologie différentielle  
28, rue Serpente  
75006 PARIS

INTRODUCTION AUX PROGRAMMES VARE, VARD, VARI (P. PERRUCHET)

Ces programmes permettent d'entrer sur disques (VARE), de décrire (VARD), et d'analyser par méthode inférentielle (VARI) des données organisées selon un plan de la forme  $\underline{S} < G > * T$  (1), sur un système APPLE II. (2).

$\underline{S}$  (pour Sujet) correspond au facteur aléatoire

G (pour Groupe) désigne ici le ou les facteurs systématiques emboîtants le facteur aléatoire. Dans la terminologie courante, qui se réfère au cas particulier où le facteur aléatoire est constitué par les sujets, on parle de facteurs "inter-individuels" et de modèle en "groupes indépendants".

T (pour traitement) désigne le ou les facteurs systématiques croisés avec le facteur aléatoire. Ces facteurs "intra-individuels" déterminent la formation de "groupes appareillés" (3).

Dans les programmes suivants, G et T peuvent être des facteurs constants, ou des facteurs composés de facteurs élémentaires croisés ; le plan doit être équilibré. Cependant, en VARI, les interactions dépassant le second ordre (entre plus de 3 variables) ne sont pas calculées.

V A R E

Le programme sollicite successivement le nombre de facteurs "inter", le nombre de facteurs "intra", puis, pour chacun des facteurs, une lettre-code et le nombre de modalités. La lettre codée servira à identifier les différents facteurs dans la série des trois programmes. L'entrée des données s'effectue en inscrivant toutes les valeurs prises par le facteur aléatoire, pour chaque combinaison de modalité des facteurs systématiques (cellule). Chaque cellule est indicée ; ainsi, l'inscription G2 T3 signifie que l'on doit rentrer les données correspondant à la 2e modalité de G et à la 3e modalité de T.

Après l'entrée d'une cellule, il est possible de corriger une ou plusieurs valeurs, en se référant aux numéros qui apparaissent entre parenthèses sur l'écran, face à chaque valeur. Lorsque les données sont correctes, elles sont enregistrées sur disque, si du moins le nombre de valeurs nouvelles est  $\geq 30$  (on peut modifier ce nombre en ligne 480).

---

(1) et dans une certaine mesure, des plans  $\underline{R} < \underline{S} < G > * T$  , cf. annexe I.

(2) Pour une adaptation à d'autres systèmes, voir annexe 2.

(3) La combinaison des facteurs inter et intra qui caractérise les plans  $\underline{S} < G > * T$  n'est pas abordée par Faverge. Elle est traitée par les anglosaxons sous le nom de "plans à mesures répétées".

Il est possible de reprendre l'entrée des données après interruption. Dans ce cas, le programme sollicite automatiquement l'entrée à la cellule immédiatement postérieure à la dernière cellule enregistrée sur le disque.

Si les données sont déjà en mémoire (par exemple dans le cas de données expérimentales directement recueillies sur ordinateur), on peut ouvrir normalement le fichier, et arrêter le programme au moment d'entrer les données (par "reset"). Il faut alors faire un programme de transfert, ouvrant le fichier par "append" ; placer, dans l'ordre d'ouverture des boucles, les facteurs inter, intra, et le fichier aléatoire.

Les données figurant dans Winer (1962, p. 324) et reproduites ci-dessous, seront reprises comme exemple dans les programmes suivants.

	Subjects	Periods : $b_1$			$b_2$			$b_3$		
		Dials : $c_1$ $c_2$ $c_3$			$c_1$	$c_2$	$c_3$	$c_1$	$c_2$	$c_3$
$a_1$	1	45	53	60	40	52	57	28	37	46
	2	35	41	50	30	37	47	25	32	41
	3	60	65	75	58	54	70	40	47	50
$a_2$	4	50	48	61	25	34	51	16	23	35
	5	42	45	55	30	37	43	22	27	37
	6	56	60	77	40	39	57	31	29	46

La formule indicée du protocole est  $S3 < A2 > * B3 * C3$ . L'enregistrement sur disque s'effectue, dans ce cas, après l'entrée de la cellule  $A2 B1 C1$  (où N atteint 30), puis à la fin des données.

#### V A R D

Après avoir rentré le nom du fichier, la formule indicée du protocole apparaît pendant quelques secondes sur l'écran, pour vérification.

Puis deux "pages" apparaissent alternativement. L'une concerne le choix des données à décrire, l'autre le type de description à effectuer sur ces données. Toutes les combinaisons possibles des éventualités proposées sur ces pages sont possibles.

Listing de la page 1 :

```

CHOIX DES DONNEES
PAS DE CHANGEMENT : RETURN
DONNEES PARTIELLES: 1
DONNEES AVEC TRANSFORMATION D'ECHELLE :2
DONNEES AVEC MOYENNAGE SUR UN FACTEUR :3
POUR UN RETOUR AUX DONNEES D'ORIGINE : 4

```

Choix 1 : la réduction peut consister à supprimer une ou plusieurs modalités d'un ou plusieurs facteurs. S'il ne reste qu'une modalité à un facteur, celui-ci n'étant plus source de variation, disparaît du protocole.

Choix 2 : on peut demander une transformation logarithmique (avec ou sans constante additionnelle), en carré, racine carrée, et inverse. Pour d'autres choix, introduire la fonction appropriée vers les lignes 720-750.

Choix 3 : le moyennage peut porter sur n'importe quel facteur. S'il porte sur le facteur aléatoire, le dernier facteur systématique devient le nouveau facteur aléatoire (voir utilisation en annexe 1). Contrairement à la réduction (1), transformation d'échelle (2) et moyennage (3) ne sont pas réalisables en VARI.II. Il convient donc d'enregistrer les données transformées si elles doivent être analysées ainsi.

Si une transformation quelconque est demandée, il est possible d'inscrire un libellé (<255 caractères, en évitant certains signes de ponctuation qui coupent le libellé) pour en indiquer la nature. Ce libellé apparaîtra sur imprimante.

Une modification demeure tant qu'une nouvelle modification n'est pas demandée (i.e. tant que l'on choisit "pas de changement" en inscrivant simplement "return"). Mais une nouvelle demande de modification (choix 1, 2 ou 3) annule automatiquement la précédente, et traite à partir des données de départ. Si l'on désire plusieurs modifications simultanées, telles que transformation d'échelle et moyennage sur un facteur, il est nécessaire d'opérer en plusieurs temps, en enregistrant sur disque les données intermédiaires. Pour revenir aux données de départ, choisir 4.

Listing de la page 2

CHOIX DE L'ANALYSE  
 DISTRIBUTION GENERALE : 1  
 DISTRIBUTION PAR FACTEUR : 2  
 MOYENNES PAR CELLULE : 3  
 MOYENNES ET ECARTS-TYPES PAR MODALITE,  
 POUR CHAQUE FACTEUR : 4  
 MOYENNES POUR LE CROISEMENT DES  
 MODALITES DE 2 FACTEURS : 5  
 POUR L'ENREGISTREMENT DES DONNEES : 6

Choix 1 : Toutes les données du protocole sont prises en compte simultanément. Les données peuvent être regroupées en 7 classes (le chiffre 7 peut être changé en affectant une autre valeur à la variable NC en ligne 870). Dans ce cas, l'effectif d'une classe est donné face à la valeur centrale de cette classe (ex. ci-dessous à gauche). Il est également possible d'établir un dénombrement des

valeurs réellement représentées. Cette opération est préférable, par exemple, lorsque le nombre de ces valeurs est petit (ex. ci-dessous à droite).

VAL. CENT.	EFFECTIFS	VALEURS	EFFECTIFS
20.4007143	3	16	1
29.2021429	9	22	1
38.0035715	14	23	1
46.805	12	25	2
55.6064286	11	27	1
64.4078572	2	28	1
73.2092858	3	29	1
		30	2
		et ...	et ...

Attention : si le nombre de valeurs différentes est  $> 50$ , et que l'on demande un dénombrement des valeurs représentées, il convient de redimensionner au préalable  $H()$  et  $R()$  en ligne 360 à une valeur au moins égale au nombre de valeurs différentes.

Choix 2 : Les distributions sont établies comme en 1, c'est-à-dire soit en 7 classes, soit à partir d'un dénombrement des valeurs existantes, mais pour chaque modalité d'un facteur déterminé.

Exemple pour le facteur B, avec regroupement en 7 classes :

#### FACTEUR B

X	MODALITES		
	1	2	3
20.400	0	0	3
29.202	0	3	6
38.003	3	6	5
46.805	5	3	4
55.606	6	5	0
64.407	2	0	0
73.209	2	1	0

Choix 3 : Les moyennes s'opèrent sur le facteur aléatoire, soit en d'autres termes, pour chaque combinaison des modalités des facteurs systématiques.

Exemple :

CELLULE	MOYENNE
A1 B1 C1	46.666
A1 B1 C2	53
A1 B1 C3	61.666
A1 B2 C1	42.666
A1 B2 C2	47.666
A1 B2 C3	58
A1 B3 C1	31
	et ...

Choix 4 : Tous les facteurs sont traités successivement et de façon automatique. Le facteur aléatoire n'est traité que sur demande de l'opérateur.

Exemple :

FACTEUR A		
MODALITES	MOYENNE	ECART-TYPE
1	47.2222222	12.7107868
2	41.3333333	14.2180816

### FACTEUR B

Choix 5 : L'opération permet, entre autre, la recherche d'éventuelles interactions de premier ordre.

Exemple :

FACTEURS	A	
	1	2
1	53.777	54.888
2	49.444	39.888
3	38.777	30.222

Attention : pour tenir en ligne, le nombre de modalités du premier facteur doit être  $< 4$  ; si les deux facteurs ont plus de 4 modalités, on devra réduire l'espacement en ligne 160.

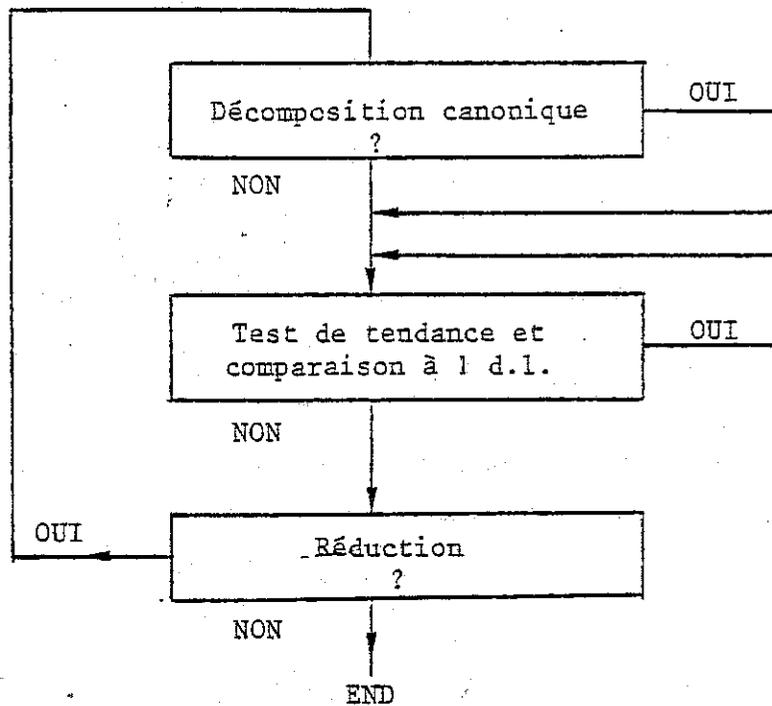
Choix 6 : Si le protocole de départ a été modifié, les nouvelles données peuvent être enregistrées sous un autre nom de fichier, pour leur traitement ultérieur en VARI. On peut également utiliser cette possibilité pour simplement répliquer un fichier sur un autre disque.

Ce choix termine l'exécution de VARD, afin d'éviter toute ambiguïté sur la nature des données analysées.

### V A R I

Ce programme permet d'exécuter la décomposition canonique associée au plan, et d'effectuer des tests de tendance (et plus généralement n'importe quelle comparaison uni-dimensionnelle) sur un facteur à partir d'un vecteur-contraste. Ces opérations peuvent être exécutées sur le protocole entier, ou sur des protocoles dérivés, obtenus par suppression d'une ou plusieurs modalités sur un ou plusieurs facteurs (Les dérivations par moyennage doivent être effectuées en VARD).

Décomposition canonique, tests de tendance et comparaisons à 1 dl, et réduction, s'ordonnent en fonction de l'organigramme suivant :



Décomposition canonique. Exemple :

PLAN D'ANALYSE : S3 ( A2 ) \* B3 \* C3

SOURCE	SC	DL	CM	F	
A	488.166	1	488.166	.751	(1/4)
B	3722.333	2	1861.166	63.388	(2/8)
C	2370.333	2	1185.166	69.823	(2/8)
AB	333	2	166.5	5.67	(2/8)
AC	50.333	2	25.166	1.907	(2/8)
BC	10.666	4	2.666	.335	(4/16)
ABC	11.333	4	2.833	.356	(4/16)
S(A)	2491.111	4	622.777		
BS(A)	234.888	8	29.361		
CS(A)	105.555	8	13.194		
BCS(A)	127.111	16	7.944		
TOTAL	9924.833	53	187.261		

POOLING DES INTERACTIONS AVEC S(A) :  
 SC: 467.555484 DL: 32 CM: 14.6111089

On notera, sur cet exemple, que les différents termes d'une interaction sont immédiatement contigus. Pour éviter toute ambiguïté lorsqu'il y a plusieurs facteurs inter-sujets, ces derniers sont séparés par le signe "\*" dans l'expression des sources de variation intra-groupe (ainsi l'absence de symbole entre 2 termes consécutifs peut toujours être remplacée par le "." d'une interaction). Pour une comparaison avec les notations anglo-saxonnes, le tableau peut être comparé à celui de Winer, p. 328 [3].

L'effet d'un facteur intra-sujet (Ex : B) est toujours évalué par rapport à l'interaction de ce facteur avec S(G) (pour l'exemple : B S(A)). Il en est de même pour les interactions intégrant au moins un facteur intra. Les F calculés correspondent donc aux F' de VAR 3 (cf. Rouanet et Lépine, 1979) [2]. Le programme fournit cependant le "pooling des interactions avec S(G)", permettant ainsi le calcul des F'' si la condition de circularité est suffisamment satisfaite. Tous les CM systématique intégrant au moins un facteur intra-sujet peuvent donc être divisé par ce CM aléatoire pour l'obtention des F'' (sur ce point, cf. Rouanet & Lépine, 1970) [1].

Tests de tendance et comparaison à 1 d.l. par entrée d'un contraste. Il suffit d'entrer, aux questions adéquates, autant de coefficients qu'il y a de modalités dans le facteur analysé.

Tests de tendance et comparaison correspondent à des choix différents, car différents CM aléatoires sont utilisés lorsque le test porte sur un facteur inter. Dans le premier cas, le CM est calculé sur l'ensemble du support du contraste, alors que dans le second cas, le CM est calculé sur les seuls termes du support différents de zéro.

Pour des tests de tendance, les coefficients des contrastes sont donnés dans la table de l'annexe 4, empruntée à Winer. Soit par exemple un test de tendance linéaire effectué sur le facteur B à 3 modalités. Il suffit d'entrer la suite : -1, 0, 1, pour obtenir :

TENDANCE LINEAIRE SUR B  
PLAN D'ANALYSE : S3 ( A2 ) \* B3 \* C3

SOURCE	SC	DL	CM	F
B	3721	1	3721	73.44 (1/4)
BA	225	1	225	4.44 (1/4)
BC	3.5	2	1.75	.287 (2/8)
BAC	1.166	2	.583	.095 (2/8)
BS(A)	282.666	4	50.666	
BCS(A)	48.666	8	6.083	

Il faut noter que la lettre B désigne ici les variations liées à la tendance linéaire de B. Par exemple, l'interaction B.A indique dans quelle mesure cette tendance est la même pour toutes les modalités de A.

Pour les comparaisons à 1 dl, on peut entrer n'importe quel contraste. Soit un facteur hypothétique à 6 modalités ; on peut demander par exemple :

Modalités :	1	2	3	4	5	6	
	0	1	-1	0	0	0	pour comparer les modalités 2 et 3
	2	-1	-1	0	0	0	pour comparer 1 à la moyenne de 2 et 3
	2	2	-1	-1	-1	-1	pour comparer les 2 premières et 4 dernières modal.

Le F obtenu correspond éventuellement au carré d'un t de Student indépendant ou apparié (s'il n'y a que 2 coefficients différents de zéro et si chacun de ces coefficients s'applique à une seule cellule). Il est cependant possible d'entrer n'importe quel contraste après demande d'un test de tendance, afin de bénéficier d'une estimation du CM aléatoire fondée sur des données plus larges (le choix inverse est possible, mais aurait peu de signification théorique).

Une comparaison à plusieurs degrés de liberté qui ne peut être réalisée par réduction du protocole (voir plus loin) nécessite autant de passages en machine qu'il y a de d.l. Il faut ensuite sommer manuellement les SC et les dl liés d'une part aux variations systématiques, d'autre part aux variations aléatoires, puis, après calcul des CM, établir le rapport F. Il est évidemment nécessaire que les contrastes soient orthogonaux, c'est-à-dire que la somme des produits des coefficients correspondants soit nulle. Dans l'exemple ci-dessus, sommer les valeurs correspondant aux 2 premiers contrastes conduit à opposer les 3 premières modalités (dans ce cas particulier, il serait cependant plus simple, pour un résultat identique, d'effectuer la décomposition canonique après avoir supprimé les modalités 4, 5 et 6 du facteur).

Réduction du protocole. La réduction peut porter sur une ou plusieurs modalités d'un ou plusieurs facteurs. Il suffit pour cela de répondre aux questions successives, comme en VARD. S'il ne reste qu'une modalité pour un facteur donné, ce facteur, n'étant plus source de variation, n'apparaît plus dans l'analyse (et en particulier dans la formule indicée du protocole qui s'imprime au sommet de chaque analyse).

Il est possible d'exécuter la décomposition canonique, un test de tendance, ou une comparaison uni-dimensionnelle, sur le protocole réduit. Voici ci-dessous le listing correspondant à une demande de tendance linéaire sur B, après suppression de la 2<sup>ème</sup> modalité de A (la 1<sup>ère</sup> ligne correspond au libellé qui peut être entré sur option, la 2<sup>ème</sup> ligne est engendrée automatiquement).

TENDANCE LINEAIRE SUR B CONDITIONNELLEMENT A A1  
 PLAN D'ANALYSE : S3 \* B3 \* C3

SOURCE	SC	DL	CM	F	
B	1058	1	1058	20.677	(1/2)
BC	2.333	2	1.166	.349	(2/4)
BS	102.333	2	51.166		
BCS	13.333	4	3.333		

Les rapports F sont toujours calculés sur le protocole réduit. Si l'on désire estimer le CM aléatoire à partir de données plus nombreuses, il est généralement nécessaire d'opérer plusieurs passages en machine.

A titre indicatif, voici comment les différents F de VAR 3 (Rouanet & Lépine, 1979) [2] peuvent être calculés, si l'on a opéré des réductions à la fois sur des facteurs inter-sujets ( $G \rightarrow G'$ ) et sur des facteurs intra-sujet ( $T \rightarrow T'$ ).

Si le test porte sur  $G'$  : le programme fournit le F1. Pour obtenir le CM permettant le calcul de F2, il faut recommencer l'analyse sur l'ensemble de G, en ne supprimant plus que les modalités du facteur T.

Si le test porte sur  $T'$  : le programme fournit le F'1. Pour obtenir F'2, faire comme ci-dessus. Pour obtenir F''1, il faut recommencer l'analyse sur l'ensemble de T, en ne supprimant plus que les modalités du facteur G, et utiliser le "pooling des interactions avec S(G)". Pour obtenir F''2, qui est le test le plus puissant, mais le moins valide, il suffit d'utiliser le "pooling des interactions" donné dans la décomposition canonique effectuée sur l'ensemble des données.

### BIBLIOGRAPHIE

- [1] ROUANET, H., LEPINE, D., Comparaisons between treatments in a repeated-measurement design ; ANOVA and multivariate methods, *British J. Math. Stat. Psychol.*, 1970, 23, 147-163.
- [2] ROUANET, H., LEPINE, D., Introduction à l'analyse des comparaisons pour le traitement des données expérimentales. *Informatique et Sciences Humaines*, réimpression 1979, 33-34, 11-125.

On trouvera dans ce document d'autres références des mêmes auteurs, sans lesquels ces programmes n'auraient pu être réalisés.

- [3] WINER, B.J., *Statistical principles in experimental design*, New York, Mc Graw-Hill, 1962.

Annexe 1 : cas où plusieurs facteurs peuvent jouer le rôle de facteur aléatoire.

VARI ne peut traité que des plans à 1 facteur aléatoire. Il est cependant possible d'entrer en VARE des plans à plusieurs facteurs aléatoires, pour les analyser et les transformer en VARD sous une forme compatible à l'analyse inférentielle. En effet, en VARE et en VARD, la classification des facteurs en systématique inter, systématique intra, et aléatoire, n'a aucune incidence. Le protocole peut donc être normalement décrit, après que des facteurs aléatoires ait été rentrés comme facteurs systématiques. Il est par contre impératif de supprimer ces facteurs en tant que sources de variation, soit en réduisant le protocole à l'une de ses modalités, soit en moyennant sur ce facteur, avant l'entrée en VARI. (le nouveau protocole doit évidemment être enregistré sur disque.).

A partir d'un même protocole de départ, il est possible de dériver, pour l'analyse, des protocoles comportant différents facteurs aléatoires. Soit un plan de la forme  $R < S < G > * T$  ( qui correspond au cas fréquent de la répétition de chaque condition élémentaire pour un sujet donné ). Ce plan peut être traité, par exemple, sous la forme  $R < G > * T$  pour chaque modalité de S, ou  $S < G > * T$  après moyennage sur R . Il convient dans ce cas de rentrer en VARE le facteur S comme le dernier facteur systématique . En suivant la procédure habituelle, il est possible de réduire le protocole à une modalité de S. Le fait de moyenner sur R conduit automatiquement à conférer à S le statut de variable aléatoire (de façon générale, le fait de supprimer le facteur aléatoire en tant que source de variation, soit par moyennage, soit par réduction à une seule modalité, conduit à conférer au dernier facteur systématique le statut de variable aléatoire). La formule indicée du plan, qui s'inscrit au début de toute analyse, permet d'authentifier le bon déroulement des opérations.

Annexe 2 : Implémentation sur d'autres systèmes utilisant le BASIC.

La seule condition restrictive concerne le dimensionnement des vecteurs, qui ne doit pas être limité ; toutes les données sont en effet placées sur un seul vecteur, avec un adressage implicite.

VARE pourra évidemment exiger des modifications assez importantes.

VARD et VARI pourront être raccourci si le "printusing" est disponible. Il convient dans ce cas de modifier en conséquence l'impression des données, et de supprimer le sous-programme facilitant l'édition des nombres sur l'APPLE II (lignes 120 à 170 en VARD, 20 à 90 en VARI).

Il apparait également que le coeur de VARI (lignes 780 à 1340) pourra être condensé de façon assez substantielle. Les capacités limitées des registres de l'APPLE II ont souvent conduit à renoncer à l'utilisation de sous-programmes, et à éclater au maximum les calculs opérés (par ex. en lignes 1130 à 1145 ou 1270 à 1276). La condensation de ces calculs est assez évidente. Le test opéré en 1500 pourra être remplacé par une boucle tournant de 1430 à 1500 (For L = 1 to N ...).

Quelques instructions spécifiques à l'APPLE :

Un "print" inclut entre "open .." "write" et "close" enregistre sur disque (RAM → disque). Un "input" inclut entre un "open" "read.." et "close.." lit le disque (disque → RAM).

Print D\$ (D\$ = CHR\$(4)) place sous le contrôle du DOS. "pr#1" renvoie à l'imprimante, "PR#0" à l'écran. Une variable suivie de % définit une variable entière (stockée sur 2 octets au lieu de 5).

ANNEXE 3 : Quelques compléments d'information.

Les données sont stockées dans un vecteur P(), de dimension N (Nb total des valeurs) L'attribution d'une donnée à une combinaison de modalités particulières s'opère à partir de l'information contenue dans 2 vecteurs de dimension NF (Nb de facteurs) ; le vecteur N() comprend le nombre de modalité, le vecteur CO(), calculé en lignes 200 et 210 de VARE, comprend le nombre de valeurs consécutives appartenant à la même modalité, ceci pour tous les facteurs ordonnés de la façon suivante : facteurs G (Nb:NI), facteurs T (Nb:NF-NI-1), facteur S.

VARI utilise presque exclusivement la méthode des contrastes (sauf pour le calcul de la SC totale en 190, et le calcul du pooling des interactions avec S(G) en lignes 1973-1979). Le programme engendre automatiquement les m contrastes orthogonaux d'une comparaison à m degrés de liberté, et procède, pour le calcul des interactions, au produit terme à terme des contrastes. Les N valeurs de chaque contraste sont utilisées immédiatement pour le calcul. Le calcul des sources de variations systématiques s'effectue au cours d'un passage unique dans le sous-programme 780-1340 (calcul successif des effets principaux, interactions de 1er et 2ème ordre), les sous-routines tournants pour les valeurs 1 (A5) à N (A6). Le calcul des sources de variations aléatoires s'effectue par cumul des valeurs recueillies en un nombre de passages dans le sous-programme 780-1340 égal au nombre de modalité de G.A5 et A6 découpant alors des blocs de valeurs égaux à CO(NI). Ainsi s'effectue la décomposition canonique (1350-1500).

Pour la comparaison uni-dimensionnelle, l'effet principal lié au contraste choisit est calculé en 1610-1640. Les autres termes (interactions et sources de variations aléatoires) sont obtenus par retour du programme de la décomposition canonique, le sous-programme pour les effets principaux calculant alors les interactions de 1er ordre, et ainsi de suite (les contrastes engendrés normalement sont systématiquement multipliés par le contraste entré par l'opérateur).

ANNEXE 4 : Contrastes pour tests de tendance

ANNEXE 4 : Contrastes pour tests de tendance

k	Polynomial	X=1	2	3	4	5	6	7	8	9	10	$\Sigma x^2$	$\lambda$
3	Linear	-1	0	1								2	1
	Quadratic	-1	-2	1								6	3
4	Linear	-3	-1	1	3							20	2
	Quadratic	1	-1	-1	1							4	1
	Cubic	-1	3	-3	1							20	1/4
5	Linear	-2	-1	0	1	2						10	1
	Quadratic	2	-1	-2	-1	2						14	1
	Cubic	-1	2	0	-2	1						10	1/4
	Quartic	-1	-4	6	-4	1						70	1/16
6	Linear	-5	-3	-1	1	3	5					70	2
	Quadratic	3	-1	-4	-4	-1	5					84	1/4
	Cubic	-5	7	4	-4	-7	5					180	1/4
	Quartic	1	-3	2	2	-3	1					28	1/16
7	Linear	-3	-2	-1	0	1	2	3				28	1
	Quadratic	5	0	-3	-4	-3	0	5				84	1
	Cubic	-1	1	1	0	-1	-1	1				6	1/4
	Quartic	3	-7	1	6	1	-7	3				154	1/16
8	Linear	-7	-5	-3	-1	1	3	5	7			168	2
	Quadratic	7	1	-3	-5	-5	-3	1	7			168	1
	Cubic	-7	5	7	3	-3	-7	-5	7			264	1/4
	Quartic	7	-13	-3	9	9	-3	-13	7			616	1/16
	Quintic	-7	23	-17	-15	15	17	-23	7			2184	1/64
9	Linear	-4	-3	-2	-1	0	1	2	3	4		60	1
	Quadratic	28	7	-8	-17	-20	-17	-8	7	28		2772	3
	Cubic	-14	7	13	9	0	-9	-13	-7	14		990	1/4
	Quartic	14	-21	-11	9	18	9	-11	-21	14		2002	1/16
	Quintic	-4	11	-4	-9	0	9	4	-11	4		468	1/36
10	Linear	-9	-7	-5	-3	-1	1	3	5	7	9	330	2
	Quadratic	6	2	-1	-3	-4	-4	-3	-1	2	6	132	1/4
	Cubic	-42	14	35	31	12	-12	-31	-35	-14	42	8580	1/4
	Quartic	18	-22	-17	3	18	18	3	-17	-22	18	2860	1/16
	Quintic	-6	14	-1	-11	-6	6	11	1	-14	6	780	1/36