

Les associations entre mensurations et leurs effets. Application à la paléontologie humaine

par André LEGUEBE

Résumé

Le diagramme de Mollison, fréquemment utilisé en paléontologie humaine pour comparer des groupes entre eux, néglige le rôle essentiel joué par le degré d'association entre les variables; il favorise en outre une confusion entre les notions de test et de distance.

L'application des méthodes de l'analyse multivariée à des échantillons d'hommes fossiles pose cependant différents problèmes en raison des spécificités des échantillons qu'il est possible de rassembler.

A un échantillon de 58 crânes du Paléolithique supérieur mesurés sous le rapport de 12 variables, est appliquée une démarche différente qui repose sur une analyse de la fiabilité de la matrice de corrélation et sur l'étude préliminaire de la position des variables dans l'espace des objets.

Une telle stratégie permet d'opérer une sélection des variables traduisant plus directement diverses caractéristiques de la morphologie; cette sélection a également pour effet de développer plusieurs espaces de référence et de situer un plus grand nombre d'objets en dépit des valeurs manquantes.

Mots-clés: diagramme de Mollison, analyse multivariée, sélection des variables, crânes, Paléolithique supérieur.

Abstract

The diagram of Mollison, often used in palaeoanthropology for the comparison of samples, completely overlooks the degree of association between the variables; it confuses the notions of test and distance.

The application of multivariate analysis procedures to samples of fossil bones, is not free from the difficulties resulting from non random and lacunar characteristics of the samples. An analysis of a sample of 58 Upper Palaeolithic skulls measured for 12 variables results in the proposal of a strategy more suited to this type of material.

Attention is focused on the confidence which can be attributed to the elements of the correlation matrix and on the importance of the vectors representative of the variables in the space of the objects. From consideration of their positions, selection is made of various sets of variables expressing more closely the patterns of morphology; such sets subtend various spaces in which a greater number of objects can be positioned in spite of missing values.

Key-words: Mollison's diagram, variable selection, Upper Palaeolithic, skulls.

Introduction

L'anthropologiste rencontre fréquemment le problème qui consiste à comparer un individu, un objet ou un échantillon à une population de référence:

a) soit pour donner une mesure de l'éloignement de l'objet du paramètre de tendance centrale correspondant de la population de référence, dans ce cas il s'agit du calcul d'une *distance*;

b) soit pour décider s'il peut accepter l'hypothèse que l'objet fait partie de la population de référence, et dans ce cas il s'agit d'un *test*.

La combinaison de résultats se rapportant à plusieurs variables différentes doit lui permettre d'obtenir, dans le cas d'une distance, une mesure plus globale et, dans le cas d'un test, de préciser son jugement. C'est cet objectif que le diagramme proposé par THEODOR MOLLISON en 1907 est sensé rencontrer. Bien qu'il viole certains principes de la statistique et qu'il risque de conduire à des déductions fausses, ce diagramme a conservé auprès de certains anthropologistes une faveur totalement imméritée. Il nous paraît donc opportun d'en faire la critique afin de mettre en évidence les points auxquels une stratégie différente doit accorder une particulière attention. LE GROS CLARK a déjà, en 1955, rappelé quelques-uns des problèmes spécifiques que pose le traitement statistique des données en paléontologie humaine.

Le diagramme de Mollison

La mesure, pour une variable déterminée, de l'éloignement d'un individu à un échantillon de référence ou *Abweichung Index*, indépendante de la grandeur de la variable, est donnée par le pourcentage que représente la différence entre la valeur observée et la moyenne, rapportée aux limites de variation de l'échantillon de référence, ces limites correspondant à celles de l'amplitude (MOLLISON, 1907). En mettant côte à côte, dans un ordre arbitraire, les résultats pour différentes variables et en adoptant une même échelle pour toutes, on obtient le « profil » de l'individu. La comparaison de divers profils d'un même individu, par rapport à plusieurs échantillons de référence est sensée permettre à l'observateur d'orienter son jugement sans toutefois conférer à celui-ci un caractère décisif.

En 1928, dans son *Lehrbuch der Anthropologie*, RUDOLF MARTIN émet des réserves au sujet de cette représentation:

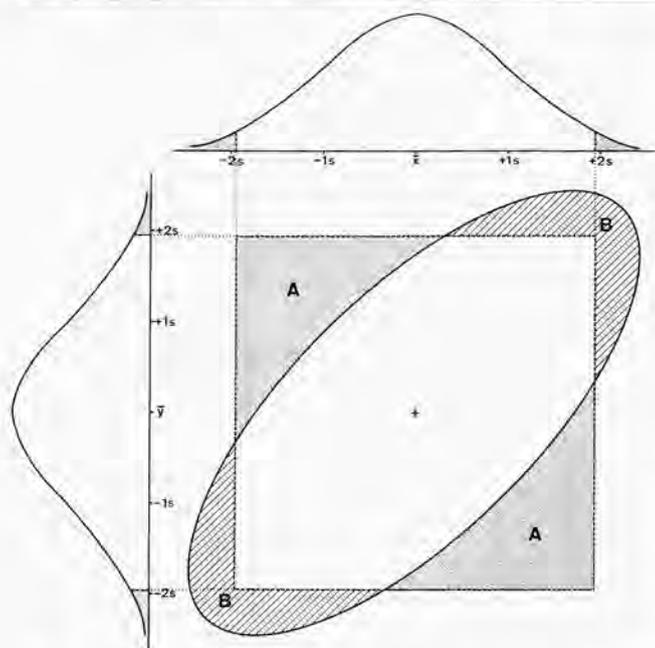


Fig. 1. – Relation entre les distributions univariées et bivariée ($r = 0.7$).

la substitution de l'écart-type à l'amplitude, paramètre critique du point de vue statistique (SIMPSON, 1941), et les modifications apportées notamment par ROTH-LUTRA (1930) et BREITINGER (1938) n'empêcheront pas GINI (1939) de condamner la méthode des profils ou des diagrammes à ordonnées jointes dans le cas de séries non ordonnées.

Au niveau univarié, l'erreur fondamentale est que la valeur que l'on porte pour caractériser l'individu n'est pas de même nature que les unités utilisées pour établir l'échelle. En effet, un individu est caractérisé par son écart normé qui est un nombre pur, alors que les unités de l'échelle sont des écarts-type qui s'expriment en mêmes unités que les variables considérées. L'extension du diagramme de MOLLISON au profil d'une population caractérisée par ses moyennes, a introduit des distorsions supplémentaires car le fait de se trouver à un nombre déterminé d'écarts-type de la moyenne de l'échantillon de référence, a une signification tout à fait différente selon qu'il s'agit d'un individu ou d'un échantillon.

Dès 1908, STUDENT avait développé un *test* destiné à éprouver le caractère significatif d'une différence entre les moyennes de deux échantillons: la différence entre deux moyennes est dépourvue de signification si on néglige de considérer la dispersion de l'échantillon comparé. Quand on est en présence de plusieurs échantillons, choisir l'un d'entre eux comme échantillon de référence introduit forcément une vision biaisée de la réalité.

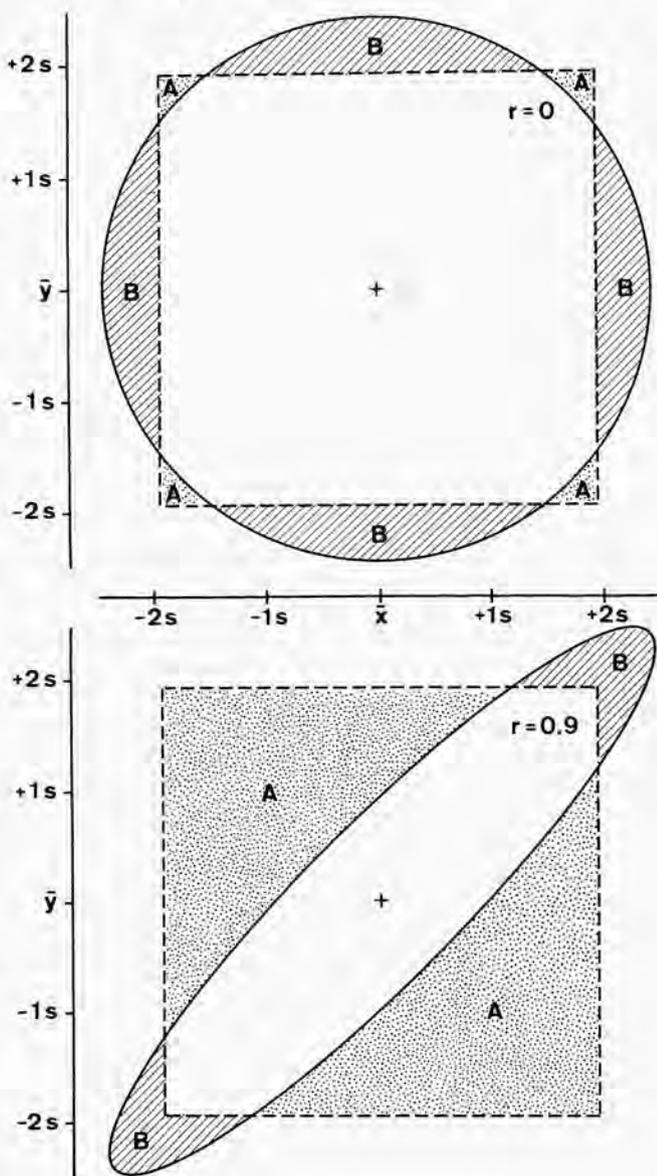
L'une des déficiences majeures du diagramme de MOLLISON réside dans le fait qu'il juxtapose simplement les variables: le tracé de profils qui négligent totalement l'importance des relations existant entre les variables peut conduire à donner une image fautive de la situation. Considérons le

cas de deux variables de moyennes \bar{x} et \bar{y} , normalement distribuées: les deux distributions se trouvent schématisées sur la Figure 1 et on a porté les limites correspondant à $-1.96s$ et $+1.96s$, entre lesquelles sont compris 95 % des objets composant la population.

Supposons ensuite qu'il existe, entre ces deux variables, un coefficient de corrélation égal à 0.7 et traçons l'ellipse qui contiendrait 95 % des objets de la population (DEFRISE, 1955; LEGUEBE, 1987). On constate, du fait de l'association existant entre les deux variables, que:

a) les objets situés dans les deux zones A, objets qui se trouvaient dans les limites de 95 % pour les deux variables considérées séparément, ne font plus partie des 95 % correspondant à la distribution bivariée;

Fig. 2. – Relation entre les distributions univariées et bivariée ($r = 0.0$ et $r = 0.9$).



b) les objets situés dans les zones B, qui étaient en dehors des limites de 95 % pour l'une ou pour les deux variables considérées séparément, sont inclus dans les limites de 95 % pour l'espace bivarié.

Cette constatation se vérifie à des degrés divers quelle que soit la valeur du coefficient de corrélation ainsi qu'on peut le constater sur la Figure 2 correspondant à des coefficients de corrélation valant respectivement 0.0 et 0.9. A première vue, comme les surfaces A et B ne contiennent évidemment qu'un nombre proportionnellement faible d'objets, on pourrait penser que ce sont les observations portant sur des objets isolés et non sur des moyennes d'échantillons qui peuvent être affectées par une telle situation.

Envisageons le cas de deux populations normalement distribuées ayant des valeurs moyennes pour l'abscisse différant de 2 écarts-type, ayant même moyenne pour l'ordonnée et dont les variables ont entre elles un coefficient de corrélation égal à 0.7 (Figure 3). Si on considère les distributions pour chaque variable séparément, il y a pour l'abscisse, une zone de recouvrement importante correspondant au secteur AB alors qu'en ordonnée, les deux distributions se superposent complètement. Si on considère l'espace bivarié, on constate que:

- il y a exclusion d'un certain nombre d'objets (zones a) et inclusion d'autres objets (zones b);
- le secteur AB commun sous le rapport de l'abscisse, se trouve disséqué en trois parties, une partie d_1 qui devient propre à la première population, une partie d_2 qui devient propre à la seconde population et une partie e qui reste commune;
- au contraire, des surfaces qui étaient, par rapport à l'abscisse, propres à l'une des deux populations, deviennent des zones de recouvrement.

Le fait de tenir compte du degré d'association entre les deux variables a donc pour effet de modifier sensiblement les limites que nous attribuons à une population et il peut affecter des zones où la densité des objets est élevée. Dans l'espace bivarié, il est essentiel de se rappeler que, si la

discrimination entre deux groupes dépend des différences entre les moyennes et des dispersions plus ou moins grandes des objets, elle est largement dépendante du degré d'association entre les variables.

Les Figures 4a et 4b montrent, pour une différence, en abscisse, de 2s entre les deux moyennes, combien l'existence d'un coefficient de corrélation entre les deux variables ($r = 0.7$) peut intensifier la discrimination entre les deux groupes. Les Figures 4c et 4d, qui correspondent aux mêmes différences entre les moyennes des deux groupes pour les abscisses et pour les ordonnées mais à des coefficients de corrélation de signes inverses (-0.7 et $+0.7$), montrent que la distribution dans l'espace de deux populations ayant mêmes moyennes et mêmes écarts-type peut être totalement différente quand les coefficients de corrélation sont de signe inverse: dans la Figure 4c, les individus proches des moyennes sont parfaitement discriminés alors que, dans la Figure 4d, ils sont inclus dans la zone où il y a recouvrement des deux populations.

L'application aveugle du diagramme de MOLLISON à des situations fondamentalement différentes a eu aussi pour effet d'ignorer la nature même des problèmes et de favoriser, dans les esprits, la confusion entre les notions de *test*, destiné à éprouver une hypothèse nulle, et de *distance* ou de *similitude* qui constitue un moyen de description. Une distance entre deux valeurs peut en effet être très grande sans être significative et inversement: il est ainsi possible de calculer des distances dont l'intérêt présente un caractère empirique alors qu'on ne dispose pas d'un test permettant de décider du caractère significatif de la différence observée.

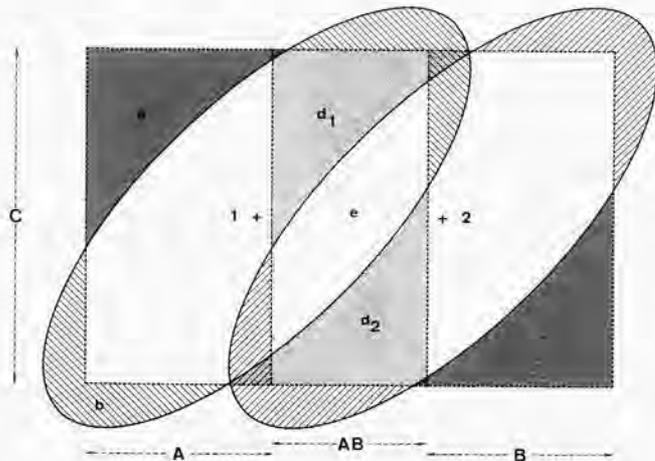
Il est essentiel d'avoir toujours présent à l'esprit que le résultat d'un test dépend des effectifs des échantillons et que, lorsqu'un test n'est pas significatif, cela ne doit pas être interprété comme l'absence d'une différence entre les échantillons ou d'une association entre variables par exemple.

Les tests univariés fournissent des réponses à des questions différentes de celles auxquelles répondent les tests multivariés, la situation multivariée n'étant pas la simple sommation des éléments univariés qui interviennent; on peut même observer des cas où le résultat d'une analyse bivariée s'avère être non significatif alors que chacune des variables considérée séparément conduit à un résultat significatif (HEALY, 1969).

Enfin, les fluctuations des mesures d'association entre variables peuvent avoir des répercussions importantes sur l'interprétation des données comme il a été constaté pour un ensemble de crânes du Paléolithique supérieur (LEGUEBE, 1988; LEGUEBE, 1989).

Dans ce travail, nous allons donc essentiellement essayer d'analyser le degré d'association pour différentes variables d'un échantillon qui n'est pas constitué de sujets prélevés au hasard et dont l'effectif est faible: nous nous limiterons au cas de variables normalement distribuées.

Fig. 3. — Effet de l'association des variables sur la discrimination entre les objets de deux échantillons.



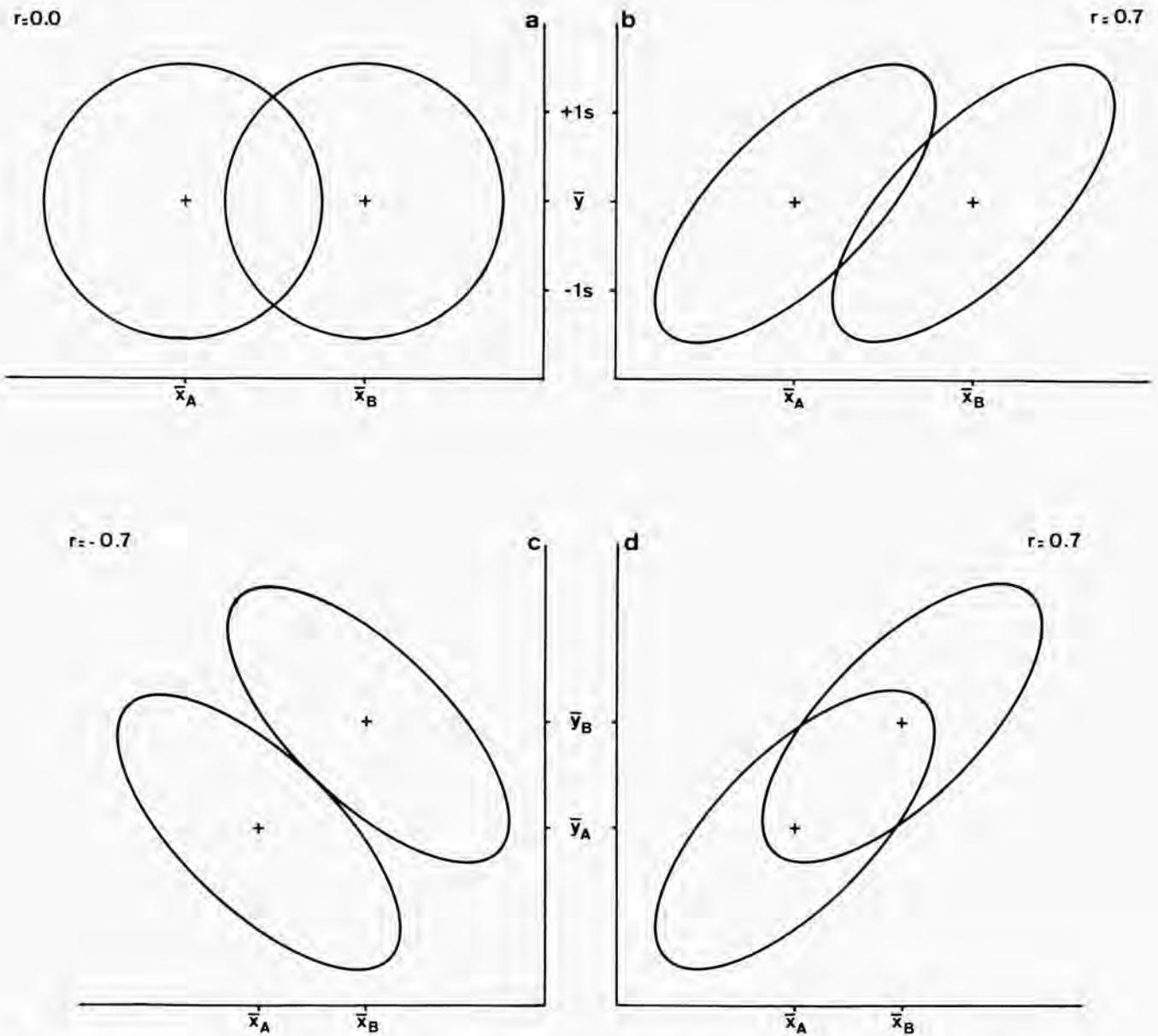


Fig. 4. – Variations de la discrimination entre deux échantillons en fonction du degré d'association de deux variables.

Matériel

L'échantillon analysé est composé de 58 crânes du Paléolithique supérieur sur lesquels douze mensurations ont été prélevées, pour autant que leur état de conservation le permettait.

Ces mensurations sont :

1. la capacité crânienne;
2. la longueur maximum;
3. la largeur maximum;
4. la largeur frontale minimum;
5. la largeur frontale maximum;
6. la hauteur basion-bregma;

7. la largeur bizygomatique;
8. la hauteur faciale supérieure;
9. la largeur orbitaire;
10. la hauteur orbitaire;
11. la largeur nasale;
12. la hauteur nasale.

L'échantillon est relativement hétérogène puisqu'il est constitué de crânes d'individus des deux sexes, ayant des âges divers, appartenant à des époques différentes et ayant des origines géographiques variées.

Ces mensurations ont été rassemblées par RIQUET (1970) et BILLY (1972).

Normalité des distributions

L'épreuve de l'hypothèse de normalité de la distribution de chacune des variables constitue une étape essentielle car son acceptation est liée à plusieurs avantages :

- a) elle valide le recours à diverses épreuves de signification;
- b) elle permet une représentation graphique de l'ensemble des éléments constituant l'échantillon;
- c) elle justifie une normalisation des mesures conduisant à l'utilisation d'une même échelle pour toutes les mensurations.

Nous avons eu recours au test de FILLIBEN (1975) dont les avantages sur les autres méthodes classiques ont été démontrés (LEGUEBE & ALBERT, 1989). Son principe consiste à comparer, au moyen d'un coefficient de corrélation, la distribution observée à une distribution théorique basée sur les rangs des objets classés par ordre de valeur croissante.

Si l'hypothèse de normalité est rejetée, différentes causes peuvent être invoquées :

- a) il s'agit d'une variable pour laquelle on a observé une distribution généralement autre que normale dans le cas d'échantillons plus fournis;
- b) l'échantillon comprend un ou plusieurs objets relativement marginaux par rapport à l'ensemble;
- c) l'échantillon comprend des objets tirés de deux ou plusieurs populations différentes.

Le fait que l'hypothèse de normalité ne soit pas rejetée n'implique pas nécessairement que l'échantillon est constitué d'objets tous tirés d'une même population, car les conditions qui doivent être remplies pour que l'hétérogénéité des populations se manifeste au niveau de la normalité de la distribution, sont assez rigoureuses. En paléontologie humaine, la faiblesse des effectifs, associée à une

variabilité relativement large des groupes que l'on peut constituer, entraîne un recouvrement important des dispersions des groupes et masque d'éventuelles tendances à la bi- ou à la plurimodalité des distributions.

Le Tableau 1 donne, pour chacune des variables, outre les effectifs, les moyennes et les écarts-type, la valeur du coefficient de corrélation du test de FILLIBEN et sa probabilité: aucune probabilité n'étant inférieure à 0.10, on peut accepter, pour toutes les variables, l'hypothèse de normalité de la distribution; les deux dernières colonnes (- et +) donnent le nombre de sujets dont les valeurs sont respectivement inférieures et supérieures à la médiane; elles permettent de constater que toutes les distributions sont relativement symétriques.

Différences entre les groupes

L'examen des graphiques associés aux tests de FILLIBEN conduit à supposer qu'il existe, pour certaines de ces variables, des différences entre les quatre groupes d'hommes aurignaciens (N = 22), de femmes aurignaciennes (N = 11), d'hommes magdaléniens (N = 14) et de femmes magdaléniennes (N = 11). Le Tableau 2 donne, pour chacun de ces quatre groupes et pour chacune des variables, l'effectif compte tenu des valeurs manquantes, la moyenne et l'écart-type.

Dans la situation qui nous occupe, le caractère significatif de l'hétérogénéité observée peut être éprouvé de deux manières différentes :

- a) soit par un test non paramétrique de KRUSKAL-WALLIS (CONOVER, 1971) qui correspond à une analyse de la variance effectuée sur les rangs, c'est-à-dire qui élimine les effets que pourrait avoir la présence d'individus marginaux;

Tableau 1. — Effectifs, moyennes, écarts-type et tests de normalité par la méthode de FILLIBEN, avec la probabilité d'obtenir une valeur r inférieure à la valeur observée.

Variables	Effectif	Moyenne	Ecart-type	r (test de FILLIBEN)	Probabilité	-	+
1. Capacité crânienne	37	1492.30	140.90	0.987	.25-.50	19	18
2. Longueur maximum	53	191.22	8.75	0.996	.95	26	27
3. Largeur maximum	56	138.79	5.87	0.995	.90	27	29
4. Largeur frontale minimum	53	97.03	5.28	0.987	.25	30	23
5. Largeur frontale maximum	40	117.80	6.58	0.985	.25-.50	20	20
6. Hauteur basion-bregma	34	135.48	7.61	0.977	.10-.25	18	16
7. Largeur bizygomatique	29	135.74	8.55	0.980	.25-.50	18	11
8. Hauteur faciale supérieure	25	67.66	5.68	0.986	.50-.75	11	14
9. Largeur orbitaire	33	41.22	3.15	0.991	.75	17	16
10. Hauteur orbitaire	34	29.19	2.02	0.976	.10-.25	15	19
11. Largeur nasale	32	25.14	2.15	0.984	.25-.50	17	15
12. Hauteur nasale	31	50.45	4.64	0.922	.75-.90	15	16

Tableau 2. — Effectifs, moyennes, écarts-type des différentes variables pour les quatre groupes de crânes du Paléolithique supérieur.

VARIABLES	AURIGNACIENS						MAGDALENIENS					
	Hommes			Femmes			Hommes			Femmes		
	N	\bar{x}	s	N	\bar{x}	s	N	\bar{x}	s	N	\bar{x}	s
Capacité crânienne	13	1580.4	141.81	6	1438.7	134.47	10	1484.5	85.65	8	1399.3	136.02
Longueur maximum	20	198.3	7.17	11	188.6	4.80	11	190.7	5.90	11	181.4	5.24
Largeur maximum	21	140.7	6.83	11	136.3	5.35	13	140.2	3.72	11	136.0	4.97
Largeur frontale minimum	20	99.3	5.96	11	97.3	3.72	12	95.7	4.08	10	93.8	4.93
Largeur frontale maximum	15	121.5	5.69	9	117.1	5.56	9	116.4	5.61	7	112.6	7.32
Hauteur basion-bregma	13	137.9	7.34	5	132.2	3.96	9	137.2	8.84	7	131.1	6.85
Largeur bizygomatique	9	139.8	8.83	7	130.6	6.65	7	140.7	6.07	6	129.9	6.71
Hauteur faciale supérieure	6	70.2	6.84	6	65.8	2.23	7	68.7	6.10	6	65.7	6.31
Largeur orbitaire	13	43.5	2.71	8	40.0	2.12	7	40.3	3.15	5	38.6	2.19
Hauteur orbitaire	13	28.4	2.22	8	29.6	2.38	7	29.6	1.79	6	30.0	0.71
Largeur nasale	10	26.3	1.97	8	26.1	1.46	8	24.8	1.83	6	22.5	1.22
Hauteur nasale	10	51.5	4.68	8	50.4	3.34	7	51.1	5.76	6	48.0	4.94

b) soit par un test d'autocomparaison (PONTIER & PERNIN, 1985) qui consiste à comparer le résultat d'une analyse classique de la variance à un critère de classification, avec les résultats des différentes analyses de la variance réalisées sur les mêmes individus regroupés différemment et au hasard un grand nombre de fois: si le résultat pour la situation observée est très fréquemment plus significatif que ceux fournis par les regroupements au hasard, on peut accepter l'hypothèse de l'existence d'une différence entre les groupes sans avoir à s'assurer que les conditions d'application de l'analyse de la variance sont effectivement

remplies.

Les résultats de ces tests figurent dans le Tableau 3 qui donne, pour chacune des variables, les effectifs des quatre groupes avec pour le test de KRUSKAL-WALLIS, la valeur du X^2 et sa probabilité, pour l'analyse de la variance le F de FISHER-SNEDECOR, les degrés de liberté, la probabilité et le rapport de corrélation, pour le test d'autocomparaison le pourcentage, sur 200 épreuves au hasard, de celles ayant fourni une valeur moindre du rapport de corrélation que celle de la situation observée, la moyenne et l'écart-type des 200 rapports de corrélation des épreuves au hasard.

Tableau 3. — Tests de KRUSKAL-WALLIS, analyses de la variance et tests d'autocomparaison pour les quatre groupes de crânes du Paléolithique supérieur.

VARIABLES	GROUPES					TEST DE KRUSKAL-WALLIS		ANALYSE DE LA VARIANCE				AUTOCOMPARAISON		
	A ♂	A ♀	M ♂	M ♀	Tot.	X^2	Probab.	D.D.L.	F	Probab.	Correl. ratio	%	Moyenne	Ecart-type
1. Capacité crânienne	13	6	10	8	37	10.65	0.01	3 33	3.92	0.01	0.262	98.0	0.084	0.0638
2. Longueur maximum	20	11	11	11	53	30.01	0.00	3 49	19.22	0.00	0.540	100.0	0.059	0.0470
3. Largeur maximum	21	11	13	11	56	6.72	0.08	3 52	2.78	0.04	0.138	93.0	0.055	0.0466
4. Largeur frontale minimum	20	11	12	10	53	5.39	0.14	3 49	3.11	0.03	0.160	99.0	0.053	0.0413
5. Largeur frontale maximum	15	9	9	7	40	9.36	0.02	3 36	3.89	0.01	0.245	100.0	0.069	0.0525
6. Hauteur basion-bregma	13	5	9	7	34	4.95	0.17	3 30	1.82	0.16	0.154	89.0	0.083	0.0573
7. Largeur bizygomatique	9	7	7	6	29	10.83	0.01	3 25	4.43	0.01	0.347	98.0	0.113	0.0818
8. Hauteur faciale supérieure	6	6	7	6	25	4.09	0.25	3 21	0.93	0.55	0.118	56.0	0.125	0.0945
9. Largeur orbitaire	13	8	7	5	33	11.91	0.00	3 29	5.74	0.00	0.372	100.0	0.091	0.0613
10. Hauteur orbitaire	13	8	7	6	34	3.96	0.26	3 30	1.20	0.32	0.107	63.0	0.096	0.0651
11. Largeur nasale	10	8	8	6	32	13.87	0.00	3 28	7.29	0.00	0.438	100.0	0.098	0.0710
12. Hauteur nasale	10	8	7	6	31	2.46	0.51	3 27	0.76	0.52	0.078	55.0	0.093	0.0764

Pour le test de KRUSKAL-WALLIS, le caractère significatif de l'hétérogénéité peut être accepté pour toutes les variables pour lesquelles la probabilité est inférieure à 0.05 soit: la capacité crânienne, la longueur maximum, la largeur frontale maximum, la largeur bizygomatique, la largeur orbitaire, la largeur nasale.

Pour le test d'autocomparaison, en adoptant la valeur de 95 % comme critère, ce sont les mêmes variables qui s'avèrent présenter une hétérogénéité des groupes significative: ce sont essentiellement, outre la capacité crânienne et la longueur maximum, les variables en relation avec la largeur du crâne et non celles en relation avec la hauteur. Les tests d'autocomparaison fournissent une information supplémentaire au niveau du rapport de corrélation qui, rappelons-le, correspond à la partie de la variabilité totale qui correspond aux différences entre groupes. Pour la capacité crânienne la valeur observée du rapport de corrélation est 3,1 fois plus grande que celle du rapport des épreuves au hasard (0.2628 / 0.0847), pour la longueur maximum 9,1 fois, pour la largeur frontale maximum 3,5 fois, pour la largeur bizygomatique 3,1 fois, pour la largeur orbitaire 4,1 fois, et pour la largeur nasale 4,5 fois.

Si l'analyse de la variance permet de conclure à la présence ou à l'absence d'une hétérogénéité entre les moyennes des groupes comparés, elle ne nous fournit, quelle que soit la procédure utilisée, aucune indication sur les groupes entre lesquels une différence existe. Le recours à des tests de STUDENT pour les groupes pris deux à deux est complètement injustifié mais une réponse peut être tirée d'un test de comparaison simultanée comme le test de GABRIEL (GABRIEL, 1964; SOKAL & ROHLF, 1981, p. 242). Les résultats de ce test sont donnés dans le Tableau 4.

Pour chaque variable, les échantillons sont rangés par ordre de moyennes croissantes et la valeur critique est calculée en fonction du seuil de signification choisi (95 % dans ce cas), des effectifs des groupes et de leur dispersion. Le test détermine le ou les ensembles maxima de groupes dont la dispersion globale n'atteint pas la valeur critique; ces groupes sont reliés par un trait; le chiffre dans la colonne de droite donne le rapport de la dispersion à la valeur critique: plus ce rapport est faible plus proches entre eux sont les éléments de l'ensemble.

Pour la grande majorité des variables, c'est le groupe des femmes magdaléniennes (Mf) qui a la moyenne la plus faible et celui des hommes aurignaciens (Ah) la moyenne la plus élevée; font exception la largeur bizygomatique et la hauteur orbitaire. Le deuxième rang, par ordre de valeurs croissantes, est occupé neuf fois sur douze par le groupe des femmes aurignaciennes (Af) et trois fois sur douze par le groupe des hommes magdaléniens (Mh). Les tests de comparaison simultanée indiquent toutefois que les chevauchements entre les groupes sont nombreux et peuvent être importants si on en juge par la valeur faible du rapport de la dispersion à sa valeur critique.

Par exemple, pour la capacité crânienne, les groupes Af, Mh, Ah constituent un ensemble par rapport à Mf et l'ensemble Mf, Af, Mh en constitue également un par opposition à Ah. Pour la largeur frontale minimum et la largeur bizygomatique, on observe des regroupements

échelonnés entre groupes voisins de la série ordonnée. Dans un certain nombre de cas, il peut y avoir un hiatus dans le regroupement (largeur frontale minimum, largeur bizygomatique, largeur orbitaire) qui est une conséquence de l'inégalité des effectifs des groupes.

Pour les variables dont l'analyse de la variance n'est pas significative (indiquée par un astérisque dans la colonne «N.S. 0.05»), l'ensemble maximum comprend nécessairement les quatre groupes. Le fait qu'aucune différence significative entre les groupes ne soit mise en évidence pour ces variables n'implique toutefois pas qu'elles soient dépourvues d'intérêt pour l'analyse multivariée puisque leur influence peut s'exercer par l'intermédiaire des associations qu'elles présentent avec d'autres variables comme il a été montré précédemment (Fig. 4a à 4d).

Association des variables

Tous les résultats d'analyses multivariées dépendent directement et essentiellement de la matrice de corrélation dont on extrait les valeurs propres et les vecteurs propres: la position des objets dans le nouvel espace défini par une analyse multivariée n'est en fait qu'une étape secondaire bien que ce soit cet aspect de l'analyse qui, d'une manière générale, ait retenu l'attention des utilisateurs.

Comme on ne dispose que d'un nombre limité d'objets, il est indispensable d'opérer une sélection des variables si on veut parvenir à dégager quelques tendances communes: une réduction du nombre de variables utilisées présente l'avantage d'augmenter le nombre des objets qu'il est possible d'inclure dans les analyses. C'est donc aux degrés d'association entre les variables qu'on doit accorder une attention particulière puisque c'est en tenant compte de leur valeur que s'opérera la sélection des variables le plus efficace.

Le degré d'association entre deux variables peut être mesuré au moyen du coefficient de corrélation de PEARSON ou du *tau* de KENDALL.

Le coefficient de corrélation de PEARSON est basé sur les valeurs primitives: quand l'effectif de l'échantillon est faible, sa valeur est relativement sensible à la présence, dans l'échantillon, d'un individu marginal (LEGUEBE, 1989).

La valeur du coefficient de corrélation de KENDALL (*tau*) est indifférente à la présence d'un sujet marginal puisque son calcul ne prend en compte que les rangs occupés par les sujets dans les deux séries de mesures: il est en outre possible d'obtenir une estimation du coefficient de corrélation de la population normale correspondante à partir du *tau* de KENDALL (KENDALL, 1970, p. 126).

En paléontologie humaine, le nombre souvent important de valeurs manquantes introduit des difficultés supplémentaires. Pour l'ensemble des variables, on peut établir une matrice de corrélation qui ne tienne compte que des objets pour lesquels on possède toutes les mensurations, ce qui a pour effet de réduire drastiquement et d'une manière qui n'est pas nécessairement aléatoire, l'effectif de l'échantillon. Dans ce cas, les estimations des moyennes et des

Tableau 4. — Tests de comparaison simultanée des groupes de crânes du Paléolithique supérieur (méthode de GABRIEL, 1964): Aurignaciens hommes (Ah), Aurignaciens femmes (Af), Magdaléniens hommes (Mh), Magdaléniens femmes (Mf).

VARIABLES	GROUPEs				% Val. crit.	N.S. 0.05
1. Capacité crânienne	Mf	Af	Mh	Ah	71. 23.	
2. Longueur maximum	Mf	Af	Mh	Ah	93. 08.	
3. Largeur maximum	Mf	Af	Mh	Ah	99.	*
4. Largeur frontale minimum	Mf	Mh	Af	Ah	99. 48. 32.	
5. Largeur frontale maximum	Mf	Mh	Af	Ah	60. 30.	
6. Hauteur basion-bregma	Mf	Af	Mh	Ah	63.	*
7. Largeur bizygomatique	Mf	Af	Ah	Mh	97. 95. 0.	
8. Hauteur faciale supérieure	Mf	Af	Mh	Ah	31.	*
9. Largeur orbitaire	Mf	Af	Mh	Ah	99. 77. 15.	
10. Hauteur orbitaire	Ah	Af	Mh	Mf	41.	*
11. Largeur nasale	Mf	Mh	Af	Ah	68. 46.	
12. Hauteur nasale	Mf	Af	Mh	Ah	26.	*

écarts-type qui sont utilisées pour le calcul des scores des individus, sont basées sur des effectifs faibles avec les conséquences que cela entraîne pratiquement (LEGUEBE, 1989).

Le calcul des coefficients de corrélation peut d'autre part utiliser, pour chaque paire de variables, le nombre maximum des valeurs présentes: dans ce cas, chaque coefficient de corrélation sera calculé sur un nombre différent d'objets.

Les valeurs des coefficients de corrélation de PEARSON pour le nombre maximum d'objets par paires de variables ($N = \max$) et pour les vecteurs complets ($N = 18$), sont donnés dans le Tableau 5; il a été montré précédemment que, pour cet échantillon, la matrice de corrélation obtenue par transformation des *tau* de KENDALL, ne conduit pas à des valeurs sensiblement différentes en raison du fait que toutes les variables sont normalement distribuées (LEGUEBE, 1988).

Tableau 5. — Matrices de corrélation (r de PEARSON pour les douze variables (pour les vecteurs complets ($N = 18$) au-dessous de la diagonale et pour le nombre maximum de paires ($N = \max$) au-dessus de la diagonale).

	1	2	3	4	5	6	7	8	9	10	11	12
1	—	.760	.420	.649	.554	.687	.650	.568	.255	-.021	.180	.528
2	.797	—	.446	.589	.597	.587	.701	.728	.577	-.056	.305	.589
3	.572	.588	—	.505	.580	.007	.734	.232	.194	-.202	-.005	.307
4	.355	.477	.676	—	.635	.310	.728	.497	.423	-.170	.225	.522
5	.659	.713	.683	.694	—	-.063	.467	.325	.157	-.347	.367	.329
6	.454	.394	-.174	-.258	.160	—	.437	.555	.162	.181	.131	.524
7	.460	.634	.625	.566	.440	.156	—	.619	.525	-.077	.236	.472
8	.446	.634	.119	.350	.391	.413	.510	—	.480	.409	.374	.872
9	-.188	.211	.044	.096	.000	-.153	.326	.294	—	.165	.431	.348
10	-.253	-.172	-.414	-.373	-.430	.195	-.232	.393	.089	—	.057	.406
11	.070	.165	-.260	.097	.166	.187	-.002	.372	.397	.165	—	.273
12	.494	.596	.229	.489	.500	.466	.441	.832	.074	.280	.365	—

Coordonnées principales

L'analyse en coordonnées principales des deux matrices de corrélation permet de repérer des groupes de variables présentant entre elles des corrélations positives et se com-

portant de manière semblable à l'égard des autres variables (GOWER, 1966; HILLS, 1969).

Les pourcentages de la variabilité totale expliqués par les différents axes et les coordonnées principales correspondantes sont donnés dans le Tableau 6.

Tableau 6. — Analyses en coordonnées principales: trois premières valeurs propres, pourcentages par rapport à la trace, coordonnées principales des variables pour les matrices de corrélation de PEARSON ($N = \max.$, trace = 6.90 et $N = 18$, trace = 7.99).

	1ère coord.		2ème coord.		3ème coord.	
	N = max.	N = 18	N = max.	N = 18	N = max.	N = 18
Valeur propre	2.14	2.87	1.31	1.64	0.95	0.94
Pourcentage	30.97	35.99	19.02	20.50	13.80	11.73
Pourcentages cumulés	—	—	49.98	56.49	63.78	68.22
Variables: scores		Rang		Rang		
1. Capacité crânienne	0.10	5	0.35	4	0.41	0.08
2. Longueur maximum	0.17	6	0.29	6	0.18	0.03
3. Largeur maximum	0.53	2	0.68	1	-0.08	-0.27
4. Largeur frontale minimum	0.38	3	0.51	3	0.04	0.03
5. Largeur frontale maximum	0.58	1	0.51	2	-0.27	0.28
6. Hauteur basion-bregma	-0.34	11	-0.41	10	0.58	0.06
7. Largeur bizygomatique	0.32	4	0.32	5	0.10	-0.30
8. Hauteur faciale supérieure	-0.30	7	-0.27	8	0.13	-0.12
9. Largeur orbitaire	-0.15	10	-0.35	9	-0.42	0.02
10. Hauteur orbitaire	-0.93	12	-0.94	12	-0.16	-0.47
11. Largeur nasale	-0.18	8	-0.54	11	-0.66	0.67
12. Hauteur nasale	-0.28	9	-0.15	7	0.16	-0.02

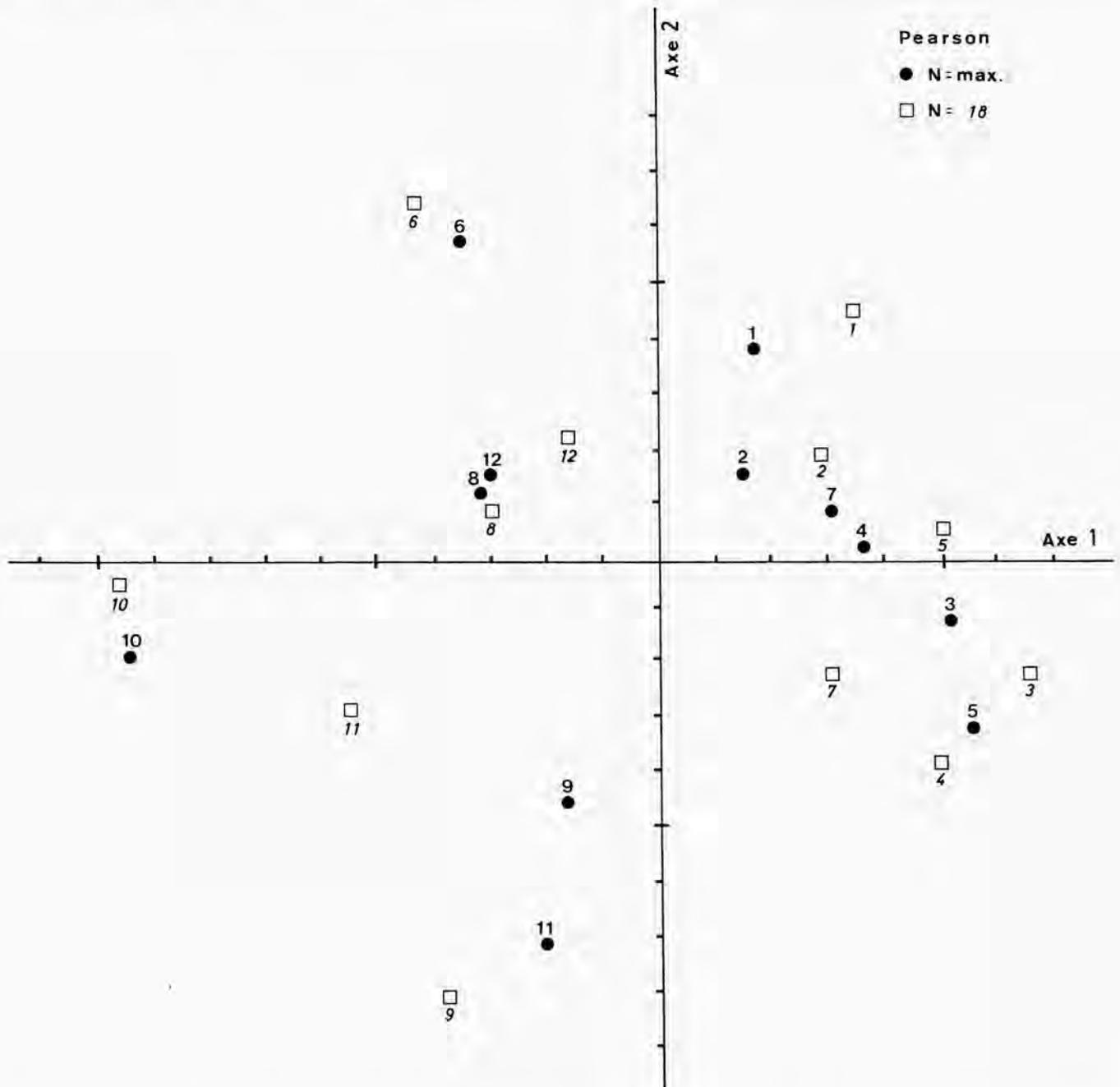


Fig. 5. – Analyse en coordonnées principales (matrices de corrélation de PEARSON pour $N = 18$ et $N = \max.$): position des extrémités des variables par rapport aux deux premiers axes.

La Figure 5 donne la position des extrémités des 12 axes variables dans l'espace des deux premières coordonnées, respectivement des matrices de corrélation pour les nombres maxima de paires ($N = \max$) et pour les vecteurs complets ($N = 18$).

Les positions des variables sont relativement semblables dans les deux cas si ce n'est qu'on observe une inversion par rapport à la deuxième coordonnée, des positions de la largeur frontale maximum (5) et de la largeur bizygomatique (7) et une modification de l'importance des rôles joués par la largeur orbitaire (9) et la largeur nasale (11).

En partant du quadrant supérieur droit et en fonction des deux premières coordonnées principales, on peut donc établir le groupement des variables suivant:

- a) la capacité crânienne (1) et la longueur maximum (2);
- b) la largeur frontale minimum (4) et la largeur frontale maximum (5) ou la largeur bizygomatique (7);
- c) la largeur maximum (3);
- d) la largeur orbitaire (9) et la largeur nasale (11);
- e) la hauteur orbitaire (10);
- f) la hauteur faciale supérieure (8) et la hauteur nasale (12);
- g) la hauteur basion-bregma (6).

Pour la troisième coordonnée (Tableau 6), la concordance entre les coefficients des variables est sensiblement plus lâche; la valeur du coefficient de la largeur maximum (3) distance cette variable des trois autres (4, 7 ou 5) dont elle se trouve plus ou moins proche pour les deux premières coordonnées.

L'espace des objets

L'analyse du rôle des variables peut aussi s'opérer à l'aide de leur projection dans l'espace des objets.

Considérons d'abord les composantes principales des douze variables, pour le nombre maximum d'objets pour chaque paire de variables ($N = \max$) et pour les vecteurs de données complets ($N = 18$).

Les pourcentages des valeurs propres par rapport à la trace des matrices figurent dans le Tableau 7. Trois composantes ayant une valeur propre supérieure à l'unité peuvent être considérées comme significatives; ensemble, elles représentent à peu près 75 % de la variabilité totale.

Si un petit nombre de valeurs propres correspondent à une large part de la variabilité, cela indique une relative homogénéité de l'échantillon mais si, au contraire, une part importante de la variabilité se distribue entre un nombre relativement grand de composantes non significatives, c'est qu'il existe une diversité difficile à caractériser soit qu'elle se rapporte à des aspects variés de la morphologie, soit que l'imprécision des mesures introduit un «bruit» qui a pour effet de rendre les tendances moins marquées.

La Figure 6 donne les projections des vecteurs «variables» dans l'espace correspondant aux deux premiers axes de l'espace des objets de l'analyse en composantes principales des matrices de corrélation pour $N = \max$ (indiquées par un trait continu) et pour $N = 18$ (extrémités indiquées par •).

Pour interpréter ce graphique, on tiendra compte des éléments suivants:

- les échelles des axes 1 et 2 sont nécessairement limitées à -1 et à $+1$;
- le cosinus de l'angle que fait un vecteur «variable» avec un axe est égal au coefficient de corrélation entre l'axe et la variable dans le plan considéré. Ainsi une variable dont l'extrémité est située à un endroit quelconque entre 0.0 et 1.0 sur l'axe 1, possède, avec cet axe, un coefficient de corrélation égal à l'unité; si cette extrémité est située entre -1.0 et 0.0, le coefficient de corrélation est égal à -1.0 . Une variable perpendiculaire à un axe présente un coefficient de corrélation nul avec cet axe, un angle de 60 degrés correspond à $r = 0.5$ et un angle de 45 degrés à $r = 0.707$. Deux vecteurs «variables» qui forment entre eux un angle petit sont fortement corrélés dans le plan considéré et inversement;
- la position de l'extrémité d'un vecteur «variable» par rapport à l'origine (0.0) indique la proportion de la variabilité totale de cette variable expliquée par les deux premières composantes; le cercle de contribution équilibrée indique quelle serait la position de l'extrémité d'un vecteur «variable» si cette variable se trouvait également répartie entre

toutes les composantes: pour une variable déterminée, le coefficient de corrélation avec un axe et la partie de sa variabilité expliquée par cet axe sont donc deux facteurs indépendants;

d) deux variables faisant entre elles un angle petit possèdent un coefficient de corrélation élevé dans le plan considéré quelle que soit la plus ou moins grande proximité des extrémités des vecteurs; inversement le coefficient de corrélation entre deux variables orthogonales est nul. Dans les quadrants supérieur et inférieur gauches de la figure 6, deux arcs de circonférence donnent les échelles des coefficients de corrélation, respectivement par rapport à l'axe 1 et par rapport à l'axe 2.

Les différences observées entre les résultats fournis par les deux matrices sont, d'une manière générale plus d'ordre quantitatif que d'ordre qualitatif et sont attribuables aux différences de composition entre les échantillons qui ont servi à calculer les matrices de corrélation. On retrouve le décalage de la largeur frontale maximum (5) observé précédemment et on observe que, pour les deux analyses, la largeur orbitaire (9) et la largeur nasale (11) se distancent l'une de l'autre, leurs pourcentages expliqués de la variabilité totale se modifiant très sensiblement. La Figure 6 établie à partir des composantes principales ne se rapporte qu'aux deux premiers axes de l'espace des objets alors que la Figure 5 donnait une représentation simplifiée de la matrice des similitudes entre les variables.

Dans le plan des deux premiers axes, se manifeste la forte corrélation entre la capacité crânienne (1) et la longueur maximum (2), la quasi-orthogonalité entre d'une part la largeur maximum (3) et d'autre part la hauteur basion-bregma (6) et la hauteur faciale supérieure (8); une situation similaire se présente entre les largeurs frontales maximum et minimum (4 et 5) et la largeur nasale (11). Les pourcentages expliqués de chacune des variables, correspondant aux composantes successives, sont donnés dans le Tableau 7: les deux dernières colonnes donnent les pourcentages de la variabilité totale des variables qui demeurent inexpliqués après la troisième composante.

Au niveau de la 1^{ère} et de la 2^e composantes, les tendances sont les mêmes dans le cas des deux matrices ($N = \max$ et $N = 18$) mais quantitativement, on observe des différences relativement importantes en particulier pour la largeur frontale maximum (5), pour la largeur orbitaire (9), et pour la largeur nasale (11).

Sur la base des résultats de l'analyse en coordonnées principales et de l'examen de l'espace des objets pour douze variables, deux séries de 6 variables sont choisies de façon à essayer de saisir des aspects différents de la variabilité. La longueur maximum (2) est utilisée dans chacune des analyses pour conserver un élément commun représentatif du format; la capacité crânienne (1) a été laissée de côté car elle est très fortement corrélée avec la longueur maximum (2) et c'est une valeur qui fait relativement souvent défaut.

La première analyse (analyse A) utilise la longueur maximum (2), la largeur maximum (3), la hauteur basion-bregma (6), la largeur bizygomatique (7), la largeur orbitaire (9) et la hauteur orbitaire (10). La seconde analyse

Tableau 7. — Analyses en composantes principales (12 variables): valeurs propres, pourcentages expliqués de la variabilité totale et pourcentages des variables expliqués par chacune des trois premières composantes.

	1er vecteur		2e vecteur		3e vecteur					
	N = max.	N = 18	N = max.	N = 18	N = max.	N = 18	N = max.	N = 18		
Valeur propre	5.64	4.95	2.02	2.46	1.21	1.55				
Pourcentage	47.01	41.29	16.84	20.47	10.14	12.90				
Pourcentage cumulé	—	—	63.85	61.76	73.99	74.66				
Variables	%	Rang	%	Rang	%	%	%	%	Non expliqué (%)	
1. Capacité crânienne	67.8	3	62.4	3	0.2	0.3	12.5	20.7	19.4	16.6
2. Longueur maximum	80.2	1	81.4	1	0.0	0.3	0.1	0.8	19.6	17.4
3. Largeur maximum	32.9	9	48.0	8	29.3	35.0	1.2	0.2	36.6	16.7
4. Largeur frontale minimum	62.9	5	51.1	6	8.2	13.8	0.0	10.5	28.9	24.7
5. Largeur frontale maximum	39.7	7	69.2	2	33.2	6.5	3.5	0.5	23.6	23.9
6. Hauteur basion-bregma	33.5	8	10.3	9	20.6	32.9	23.6	35.4	22.3	21.4
7. Largeur bizygomatique	73.0	2	56.7	4	3.9	1.4	0.3	4.7	22.8	37.2
8. Hauteur faciale supérieure	67.4	4	48.8	7	19.4	36.1	0.3	1.3	12.9	13.2
9. Largeur orbitaire	32.7	10	2.8	12	2.8	5.3	28.1	66.1	36.3	15.8
10. Hauteur orbitaire	0.3	12	4.6	10	66.8	59.1	0.5	0.2	32.3	16.2
11. Largeur nasale	16.2	11	4.6	11	1.4	30.9	51.2	14.4	31.1	51.1
12. Hauteur nasale	57.3	6	55.5	5	16.3	24.1	0.2	0.1	26.1	20.3

Tableau 8. — Analyses en composantes principales pour deux ensembles de six variables: valeurs propres, pourcentages expliqués de la variabilité totale et pourcentages des variables expliqués par les trois premières composantes.

	Analyse A 6 variables (2, 3, 6, 7, 9, 10) N = 25			Analyse B 6 variables (2, 4, 5, 8, 11, 12) N = 24		
	VALEURS PROPRES ET POURCENTAGES					
1ère composante	2.85	47.42		3.19	53.15	
2ème composante	1.32	22.04		0.95	15.86	
3ème composante	0.86	14.40		0.86	14.38	
Total ($\lambda > 1$)		69.46			53.15	
VARIABLES	COMPOSANTES					
	1	2	3	1	2	3
1. Capacité crânienne						
2. Longueur maximum	80.3	1.0	2.3	64.4	2.7	4.3
3. Largeur maximum	43.4	32.4	2.9			
4. Largeur frontale minimum				43.5	24.0	1.3
5. Largeur frontale maximum				56.9	21.9	4.5
6. Hauteur basion-bregma	31.7	27.1	38.0			
7. Largeur bizygomatique	85.5	2.7	0.1			
8. Hauteur faciale supérieure				64.4	19.2	9.3
9. Largeur orbitaire	43.5	5.9	27.8			
10. Hauteur orbitaire	0.1	63.1	15.3			
11. Largeur nasale				19.6	16.8	62.5
12. Hauteur nasale				70.2	10.7	4.4

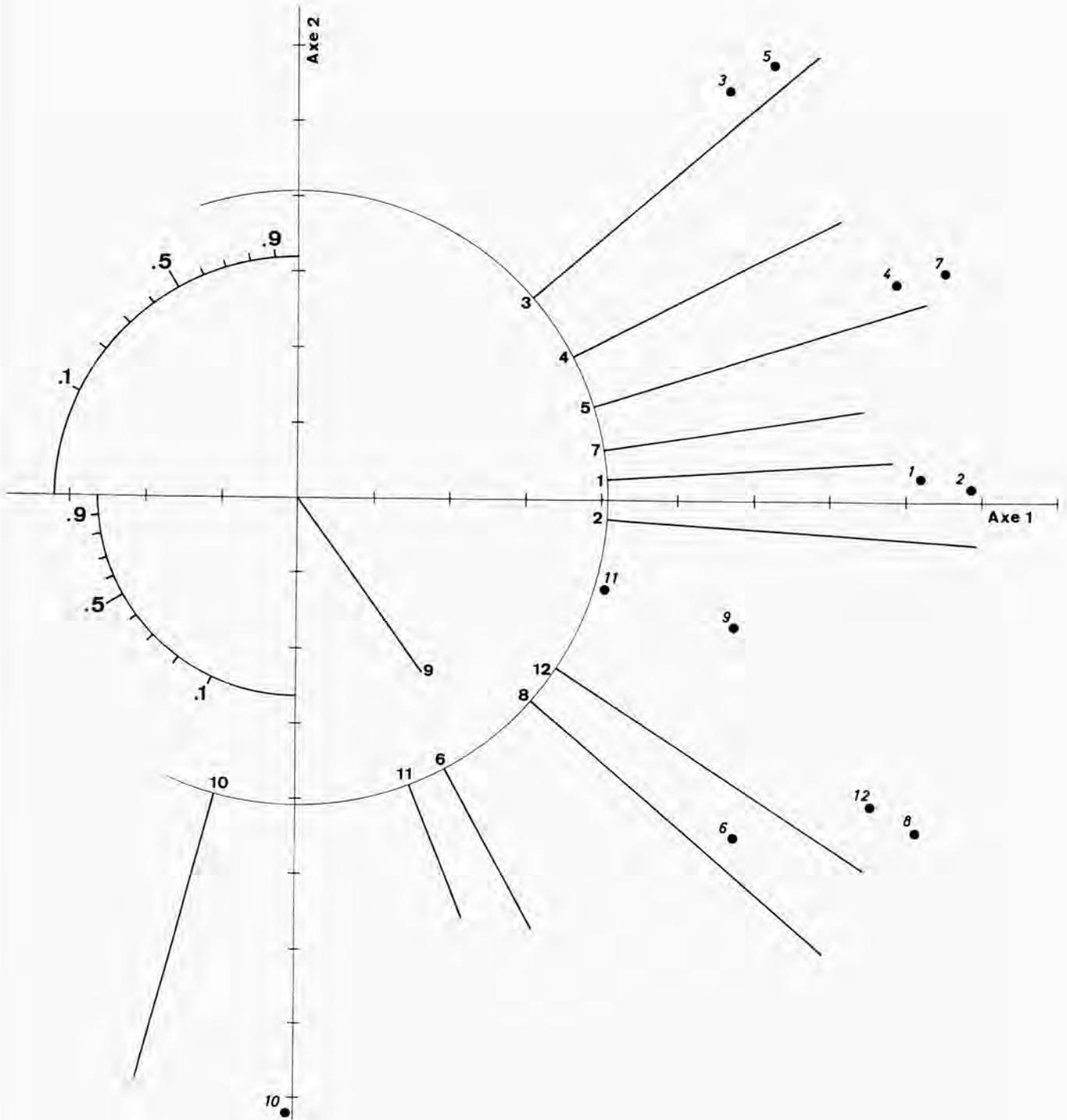


Fig. 6. - Analyse en composantes principales sur les matrices de corrélation pour les 12 variables : position des vecteurs « variables » dans l'espace des 2 premiers axes (*r* de PEARSON avec $N = \max.$ et avec $N = 18$, chiffres en italiques).

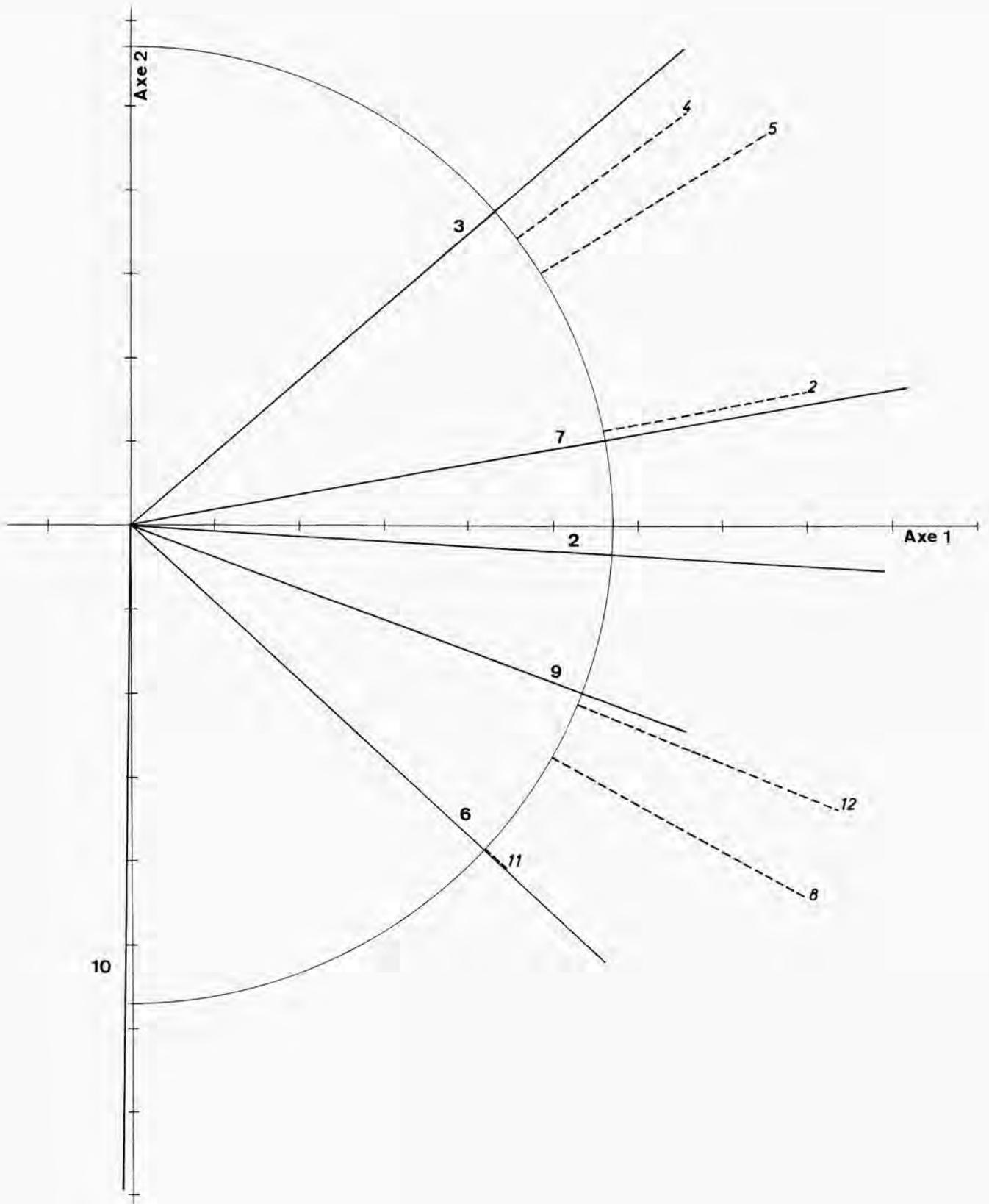


Fig. 7. — Analyse en composantes principales sur les matrices de corrélation (r de PEARSON) pour deux séries de 6 variables (Analyse A: var. 2, 3, 6, 7, 9, 10, $N = 25$; Analyse B: var. 2, 4, 5, 8, 11, 12, $N = 24$): position des vecteurs «variables» dans l'espace des 2 premiers axes.

(analyse B) tient compte de la longueur maximum (2), la largeur frontale minimum (4), la largeur frontale maximum (5), la hauteur faciale supérieure (8), la largeur nasale (11) et la hauteur nasale (12); 25 crânes sont inclus dans la première analyse, et la seconde utilise les données de 24 crânes: 19 crânes sont communs avec la première. Le calcul de la matrice de corrélation est réalisé, dans chaque cas, à partir des vecteurs complets pour les 6 variables. Les pourcentages des valeurs propres rapportées à la trace de la matrice (trace = 6.0) figurent dans le Tableau 8. Le nombre de valeurs propres supérieures à l'unité est moindre que dans le cas de douze variables. Toutefois les positions des vecteurs «variables» résultant de ces analyses sont relativement semblables à celles obtenues précédemment pour les douze variables. L'utilisation de différents ensembles de variables ayant peu d'éléments en commun, aboutit à une décomposition de la variabilité globale en segments représentatifs d'aspects divers de la morphologie.

Considérons les pourcentages des variables expliquées pour chacune des composantes des analyses A et B (Tableau 8): ils sont approximativement du même ordre de grandeur que ceux obtenus dans l'analyse basée sur douze variables mais des variations importantes peuvent être enregistrées pour une variable particulière selon le jeu de variables dans lequel elle se trouve comprise comme le montre l'exemple de la longueur maximum (2).

De nombreuses combinaisons différentes des variables sont évidemment possibles et leur intérêt sera déterminé en fonction de ce qu'elles apportent à la qualification des objets composant l'échantillon: le jeu des variables à utiliser pourra donc varier selon les échantillons à analyser.

Conclusion

PEARSON, dès 1904, soulignait:

«To the statistician, nothing is more obvious and intelligible than the independence of the correlations, variations and means of characters. A knowledge of any one of the three involves no knowledge of the other two.»

L'étude de la variabilité, variable par variable, ne nous fournit qu'une vue limitée des relations entre les objets composant l'échantillon. Les mensurations utilisées actuellement sont en fait le résultat d'une assez longue tradition visant à traduire divers aspects de la morphologie, à assurer une bonne reproductibilité des mesures et à offrir une possibilité de comparaison entre les résultats de différents observateurs.

L'application de l'analyse multivariée aux données de la paléontologie humaine réclame qu'une attention toute particulière soit accordée aux relations existant entre les variables.

L'introduction dans une même analyse d'un trop grand nombre de variables a pour effet de rendre l'interprétation des résultats peu significative ou même floue et il est donc essentiel d'opérer un choix parmi celles dont on dispose. Les méthodes de sélection des variables généralement utilisées font uniquement appel à des critères statistiques qui peuvent ne pas être adaptés au traitement d'échantillons tels que ceux que nous pouvons constituer en paléontologie humaine.

L'analyse de la structure de la matrice de corrélation par la méthode des coordonnées principales et du rôle des variables au moyen de l'espace des objets que fournit l'analyse en composantes principales, et ceci avant d'étudier la distribution des objets eux-mêmes dans un nouvel espace, présente de nombreux avantages.

Elle permet:

- a) de sélectionner des variables qui, compte tenu des associations existant entre elles, traduisent plus directement certains caractères de la morphologie;
- b) d'éviter, au moins en partie, une réduction de l'effectif résultant de l'absence de certaines valeurs;
- c) de tenir compte des particularités de l'échantillon; par exemple, pour le matériel paléontologique, l'état de conservation des crânes ne permet pas fréquemment de mesurer la capacité crânienne ou la hauteur basion-bregma;
- d) de traiter des objets qu'on ne peut inclure dans une même analyse parce que les valeurs manquantes concernent des mensurations différentes; en rapportant ces objets appartenant à un même ensemble commun, à différentes combinaisons de variables, la possibilité est offerte d'analyser un nombre maximum d'objets composant l'échantillon.

La sélection de variables opérée avant l'application des méthodes d'analyse des profils tenant compte des relations entre les variables (GEISSER & GREENHOUSE, 1958; GREENHOUSE & GEISSER, 1959; MORRISON, 1967, p. 145 et p. 190) aura pour effet soit de répondre plus directement aux questions du chercheur, soit de déboucher sur une description plus synthétique de la structure des variations observées. Enfin, si on envisage d'opérer sur différents jeux de variables et de comparer les résultats, il deviendra indispensable de préciser dans quelle mesure les différences observées concernent le format ou la forme, problème qui n'a pas encore reçu de solution entièrement satisfaisante.

Une analyse préliminaire des variables, combinée avec un examen critique des coefficients de corrélation (LEGUEBE, 1988) et le calcul d'indices d'atypisme (ALBERT & LEGUEBE, 1989) permet d'envisager le développement d'une stratégie nouvelle dont le résultat aboutit à une représentation plus diversifiée des rapports existant entre les objets et à caractériser de manière plus ponctuelle ce qui les rapproche et ce qui les distingue.

Index bibliographique

- ALBERT, A. & LEGUEBE, A., 1989. Indices d'atypisme en paléontologie humaine. *Zeitschrift für Morphologie und Anthropologie*, 77 (3) : 273-286.
- BILLY, G., 1970. Définition du type de Cro-Magnon «sensu stricto». In CAMPS, G. & OLIVIER, G., L'homme de Cro-Magnon. Anthropologie et Archéologie. Paris, Arts et Métiers graphiques, pp. 23-35.
- BILLY, G., 1972. L'évolution humaine au Paléolithique supérieur. *Homo*, 23 (1/2) : 2-12.
- BREITINGER, E., 1938. Zur Differentialdiagnose zwischen nordischen und mittelländischen Schädeln. *Verhandlungen der deutschen Gesellschaft für Rassenforschung*, 9 : 113-122.
- CONOVER, W.J., 1971. Practical nonparametric statistics. New York, Wiley and Sons, XII, 462 p.
- DEFRISE-GUSSENHOVEN, E., 1955. Ellipses équiprobables et taux d'éloignement en biométrie. *Bulletin de l'Institut royal des Sciences naturelles de Belgique*, 21 (26) : 31 p.
- FILLIBEN, J.J., 1975. The probability plot correlation coefficient test for normality. *Technometrics*, 17 : 111-117.
- GABRIEL, K.R., 1964. A procedure for testing the homogeneity of all sets of means in analysis of variance. *Biometrics*, 20 : 459-477.
- GEISSER, S. & GREENHOUSE, S.W., 1958. An extension of Box's results on the use of F distribution in multivariate analysis. *Annals of mathematical Statistics*, 29 : 885-891.
- GINI, C., 1939. Sur la méthode des profils. *Revue de l'Institut international de Statistique*, 7 : 1-31, 225-246.
- GOWER, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53 : 325-338.
- GREENHOUSE, S.W. & GEISSER, S., 1959. On methods of analysis of profile data. *Psychometrika*, 24 : 95-112.
- HEALY, M.J.R., 1969. Rao's paradox concerning the multivariate tests of significance. *Biometrics*, 25 : 411-413.
- HILLS, M., 1969. On looking at large correlation matrices. *Biometrika*, 56 (2) : 249-253.
- HOWELLS, W.W., 1969. Criteria for selection of osteometric dimensions. *American Journal of physical Anthropology*, 30 : 451-458.
- KENDALL, M.G., 1970. Rank correlation methods. London, Griffin and Co, 4th ed., 202 p.
- LE GROS CLARK, W.E., 1955. The fossil evidence for human evolution. An introduction to the study of paleoanthropology. Univ. Chicago Press, X, 181 p., 20 fig.
- LEGUEBE, A., 1986. Méthodes biométriques. In FEREMBACH, D., SUSANNE, CH., & CHAMLA, M.-CL. (Ed.), L'homme, son évolution, sa diversité. Paris, Doin, pp. 43-63.
- LEGUEBE, A., 1988. Les fluctuations des coefficients de corrélation en paléontologie humaine. *Bulletin de la Société royale belge d'Anthropologie et de Préhistoire*, 99 (sous presse).
- LEGUEBE, A., 1989. Analyse biométrique des données en paléontologie humaine. *L'Anthropologie*, Paris (sous presse).
- LEGUEBE, A. & ALBERT, A., 1989. Test de normalité graphique en paléontologie humaine. *Zeitschrift für Morphologie und Anthropologie*, 77 (3) : 259-271.
- MARTIN, R., 1928. Lehrbuch der Anthropologie in systematischer Darstellung. Jena, Gustav Fischer Verlag, 2. Aufl.
- MOLLISON, T., 1907. Die Maori in ihren Beziehungen zu verschiedenen benachbarten Gruppen. *Korrespondenz-Blatt der deutschen Gesellschaft für Anthropologie*, 38 : 147-152.
- MOLLISON, T., 1910. Die Körperproportionen der Primaten. *Morphologisches Jahrbuch*, 42 : 79-304.
- MORRISON, D.F., 1967. Multivariate statistical methods. New York, McGraw-Hill Co., 338 p.
- PEARSON, K., 1904. A Mendelian's view of the law of ancestral inheritance. *Biometrika*, 3 : 109-112.
- PONTIER, J. & PERNIN, M.-O., 1985. Les tests d'autocomparaison. Applications à l'anthropologie. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, sér. XIV, 2 : 259-276.
- RIQUET, R., 1970. La race de Cro-Magnon : abus de langage ou réalité objective? In CAMPS, G. & OLIVIER, G., L'homme de Cro-Magnon. Anthropologie et Archéologie. Paris, Arts et Métiers graphiques, pp. 37-58.
- ROTH-LUTRA, K., 1930. Das Qualifikationsdiagramm und das Wahrscheinlichkeitsdiagramm. *Verhandlungen der deutschen Gesellschaft für physische Anthropologie*, 4 : 31-57.
- SIMPSON, G.G., 1941. Range as a zoological character. *American Journal of Science*, 239 (11) : 785-804.
- SOKAL, R.R. & ROHLF, F.J., 1981. Biometry. The principles and practice of statistics in biological research. San Francisco, Freeman and Co, 2d ed., 859 p.
- STUDENT, 1908. Probable error of a mean. *Biometrika*, 6 : 1-13.

André LEGUEBE
 Institut royal des Sciences naturelles
 de Belgique
 Section d'Anthropologie
 et de Préhistoire,
 rue Vautier, 29
 B-1040 Bruxelles.