BIOMETRIE-PRAXIMETRIE — VI, 1, 36-47, 1965

LA RECHERCHE DES CONSTELLATIONS DE GROUPES A PARTIR DES DISTANCES GENERALISEES D^2 DE MAHALANOBIS

par

Raymond VAN DEN DRIESSCHE

Office de la Recherche Scientifique et Technique Outre-Mer, Bondy (Seine)

Lorsque les unités d'échantillonnage d'une enquête se classent dans une ou plusieurs dizaines de *groupes* et sont caractérisées par de nombreuses *variables*, il est essentiel d'évaluer les *distances* entre ces groupes.

Ce qui est visible, pour deux variables centrées réduites, dans un espace à deux dimensions renfermant quatre groupes

.G3

G1.

.G4

.G2

où il y a six distances intergroupes,

se conçoit, pour une dizaine de variables, dans un espace à dix dimensions, par l'estimation du paramètre statistique D^2 de Mahalanobis (1925, 1927, 1930, 1936, 1937, 1948) encore appelé *distance généralisée*. Si $x_{p,k,t}$ est une donnée de la variable p déterminée sur l'unité d'échantillonnage t, dans le groupe k (le nombre de variables étant constant pour toutes les unités d'échantillonnage), et si n_k et n désignent, respectivement, le nombre d'unités d'échantillonnage par groupe et le nombre de groupes, la moyenne et l'écart-type intra-groupe

$$m_{p,k} = \Sigma x_{p,k,t}/n_k \qquad s_{p,k} = [\Sigma(x_{p,k,t} - m_{p,k})^2/(n_k - 1)]^{1/2}$$

et les paramètres correspondants pour l'ensemble des groupes

$$m_p = \Sigma m_{p,k}/n \qquad s_p = [\Sigma(n_k - 1)s^2_{p,k}/(\Sigma n_k - n)]^{1/2}$$

donnent les moyennes centrées réduites intra-groupe

$$x_{p,k} = (m_{p,k} - m_p)/s_p$$

Celles-ci, de concert avec les corrélations totales d'ensemble,

$$r_{p,p-1} = [\Sigma(\Sigma x_{p,k,t} x_{p-1,k,t} - n_k m_{p,k} m_{p-1,k})/(\Sigma n_k - n)]/s_p s_{p-1}$$

sont soumises à un processus de transformation (Rao, 1952, p. 346) donnant des moyennes $y_{p,k}$ non corrélées et de variance unitaire. Les $(n-1)n/2$ distances généralisées sont les sommes de carrés

$$D^2_{k,k-1} = (y_{p,k} - y_{p,k-1})^2 + (y_{p-1,k} - y_{p-1,k-1})^2 + \ldots + (y_{1,k} - y_{1,k-1})^2$$

Tiré, comme exemple numérique, d'une série d'analyses par les D^2 entamée début 1964, le tableau 1 est obtenu entre 11 groupes de sols d'Afrique de l'Ouest à partir de 11 variables analytiques déterminées sur l'horizon de surface de 1008 profils. Le résultat de ces travaux paraîtra dans les *Cahiers de Pédologie de l'O.R.S.T.O.M.*

Après l'établissement d'un tableau de D^2 intergroupes, il y a lieu de se demander à quel point ces D^2 différencient les groupes étudiés, et si certains groupes très proches, du fait des D^2 faibles qui les relient, ne forment pas des ensembles indissociables, ou des ensembles que les variables utilisées ne parviennent pas à dissocier, et qui portent le nom de *constellations*. Déterminer les variables les plus puissantes dans la dissociation et estimer des affinités entre certains groupes, sous-groupes, familles ou séries de la classification des sols, constituent des objectifs intéressants.

L'examen attentif d'un tableau de distances généralisées D^2 obtenues sur une dizaine de groupes permet, à lui seul, de déceler certaines constellations. Pour quelques dizaines de groupes cela devient laborieux et il est de toute façon plus rigoureux et plus rapide de procéder *par calcul.*

Des méthodes sont décrites par Rao (*op. cit.*, p. 363): « Un processus simple avancé par K.D. Tocher consiste à prendre pour point de départ deux groupes étroitement associés et à trouver le troisième groupe qui présente, par rapport aux deux premiers, le D^2 moyen le plus faible. Un quatrième groupe est choisi de la même façon pour son D^2 moyen le plus faible par rapport aux trois premiers, et ainsi de suite. Si, à une étape quelconque, le D^2 moyen d'un groupe paraît élevé comparé aux précédents, le groupe en question ne convient pas; il est alors considéré comme se trouvant à l'extérieur de la constellation. Les groupes de la première constellation étant ensuite négligés, on traite les autres groupes de façon similaire. » « Il est également utile de calculer le changement de D^2 moyen intra-constellation provoqué par l'introduction d'un groupe additionnel. Si le changement est appréciable, il faut considérer le groupe nouvellement ajouté comme étant hors de la constellation. » Comme on le voit, ces méthodes font appel au jugement et il semble bien difficile de décider quelle valeur, ou rapport de valeurs, doit être considéré comme *appréciable* ou *élevé*. Une remarque formulée à ce sujet par Rao (*op. cit.*, p. 362) a retenu toute l'attention: « Des règles impératives pour la recherche de constellations ne peuvent être énoncées, car la constellation n'est pas un terme bien défini. Que deux groupes d'une même constellation doivent avoir, en moyenne, un D^2 inférieur à celui liant deux groupes appartenant à des constellations différentes semble être le seul critère. »

Une approche numérique du problème de la recherche des constellations est tentée. Des conditions précises sont choisies; elles permettent de décider si des groupes entrent ou non dans des constellations, sans faire appel aux adjectifs qualificatifs susmentionnés.

La première constellation

Les D^2 sont présentés, par groupe, dans l'ordre croissant. Il y a, au total, $(n-1)n/2$ distances entre les n groupes, mais la présentation matricielle n'est guère pratique. La différence entre le D^2 maximal et le D^2 minimal est appelée étendue. Le D^2 le plus faible du tableau relie deux groupes dont l'appartenance à une première constellation n'est mise en doute que lorsque ce D^2 dépasse l'étendue.

La lecture des D^2 figurant en tête des colonnes 1,45 3,19 1,45 1,29 ... du tableau 1 désigne le D^2 minimal 1,28 = G10 à G11. En bas de colonnes, on repère le D^2 maximal 32,43. L'étendue 32,43 − 1,28 = 31,15 dépasse 1,28 et G10 + G11 = C1 constitue l'ébauche de la première constellation; cette étape figure en tête du tableau 2.

L'adjonction d'un troisième groupe se fait sur la base d'un accroissement minimal du premier D^2. Ce premier D^2 ne peut être accru, pour un troisième groupe, que de la somme de deux D^2, mais il existe $n-2$

Tableau 1. — Distances généralisées intergroupes D^2

Du groupe G1 aux groupes		Du groupe G2 aux groupes		Du groupe G3 aux groupes		Du groupe G4 aux groupes		Du groupe G5 aux groupes		Du groupe G6 aux groupes		Du groupe G7 aux groupes		Du groupe G8 aux groupes		Du groupe G9 aux groupes		Du groupe G10 aux groupes		Du groupe G11 aux groupes	
G3	1,45	G11	3,19	G1	1,45	G5	1,29	G4	1,29	G1	3,36	G6	6,61	G6	14,09	G10	5,69	G11	1,28	G10	1,28
G6	3,36	G10	4,07	G5	6,26	G1	5,64	G1	3,87	G11	4,29	G11	10,14	G10	14,83	G11	7,27	G2	4,07	G2	3,19
G11	3,62	G1	7,59	G4	6,57	G3	6,57	G6	4,49	G5	4,49	G1	10,25	G9	14,93	G2	10,63	G1	4,91	G1	3,62
G5	3,87	G6	9,68	G6	7,13	G6	8,41	G3	6,26	G10	5,02	G5	10,53	G11	15,06	G6	14,28	G6	5,02	G6	4,29
G10	4,91	G9	10,63	G11	8,29	G11	11,79	G11	7,95	G7	6,61	G10	10,62	G7	17,37	G8	14,93	G9	5,69	G9	7,27
G4	5,64	G3	11,10	G10	9,40	G10	12,13	G10	8,34	G3	7,13	G2	11,56	G2	21,60	G1	16,29	G5	8,34	G5	7,95
G2	7,59	G7	11,56	G2	11,10	G7	13,53	G7	10,53	G4	8,41	G3	11,58	G1	22,84	G7	16,68	G3	9,40	G3	8,29
G7	10,25	G5	15,27	G7	11,58	G2	18,63	G2	15,27	G2	9,68	G4	13,53	G5	25,96	G5	20,56	G7	10,62	G7	10,14
G9	16,29	G4	18,63	G9	22,34	G9	25,24	G9	20,56	G8	14,09	G9	16,68	G3	29,31	G3	22,34	G4	12,13	G4	11,79
G8	22,84	G8	21,60	G8	29,31	G8	32,43	G8	25,96	G9	14,28	G8	17,37	G4	32,43	G4	25,24	G8	14,83	G8	15,06

Tableau 2. — Etapes de la détermination des constellations *C*

Groupe	Accroissement de D^2	Somme de D^2	Nombre de D^2	Moyenne de D^2	Constellation	Conditions
		1,28	1	1,28	C1=G10+G11	31,15 > 1,28
G1	8,53					
G2	7,26	8,54	3	2,85	C1=G10+G11+G2	G2G1 = 7,59 > 2,85
G3	17,69					G11G1 = 3,62 > 2,85
G4	23,92					G10G1 = 4,91 > 2,85
G5	16,29					
G6	9,31					
G7	20,76					
G8	29,89					
G9	12,96					
G1	16,12	24,66	6	4,11	C1=G10+G11+G2+G1	G1G3 = 1,45 < 4,11 (*) G1 à l'extérieur de la constellation
G3	28,79					
G4	42,55					
G5	31,56					
G6	18,99					
G7	32,32					
G8	51,49					
G9	23,59					
					2ᵉ constellation = C2	
		1,29	1	1,29	C2=G4+G5	C2C1 = 12,35 > 1,29; 2,85
G1	9,51	10,80	3	3,60	C2=G4+G5+G1	G1G3 = 1,45 < 3,60 (*) G1 à l'extérieur de la constellation
G3	12,83					
G6	12,90					
G7	24,06					
G8	58,39					
G9	45,80					

Groupe	Accroissement de D^2	Somme de D^2	Nombre de D^2	Moyenne de D^2	Constellation	Conditions
					3e constellation = C3	
		1,45	1	1,45	C3=G1+G3	C3C1 = 7,49 > 1,45; 2,85; 1,29
						C3C2 = 5,59 > 1,45; 2,85; 1,29
						C2C1 = 12,35 > 1,45; 2,85; 1,29
G6	10,49	11,94	3	3,98	C3=G1+G3+G6	G6G7 = 6,61 > 3,98
G7	21,83					G3G7 = 11,58 > 3,98
G8	52,15					G1G7 = 10,25 > 3,98
G9	38,63					C3C1 = 7,10 > 3,98; 2,85; 1,29
						C3C2 = 5,87 > 3,98; 2,85; 1,29
						C2C1 = 12,35 > 3,98; 2,85; 1,29
G7	28,44	40,38	6	6,73	C3=G1+G3+G6+G7	G7G9 = 16,68 > 6,73
G8	66,24					G6G8 = 14,09 > 6,73
G9	52,91					G3G9 = 22,34 > 6,73
						G1G9 = 16,29 > 6,73
						C3C1 = 8,02 > 6,73; 2,85; 1,29
						C3C2 = 7,41 > 6,73; 2,85; 1,29
						C2C1 = 12,35 > 6,73; 2,85; 1,29
G8	83,61					
G9	69,59	109,97	10	11,00	C3=G1+G3+G6+G7+G9	G9G8 = 14,93 > 11,00
						G7G8 = 17,37 > 11,00
						G6G8 = 14,09 > 11,00
						G3G8 = 29,31 > 11,00
						G1G8 = 22,84 > 11,00
						C3C1 = 7,99 < 11,00 (**)
						G9 à l'extérieur de la constellation
					4e constellation = C4	
		14,93	1	14,93	C4=G8+G9	C4C1 = 12,51 < 14,93 (**)
						Pas de 4e constellation

(*) La condition D^2 intra-extra-constellation > $\bar{D}^2$ intra-constellation n'est plus remplie.
(**) La condition $\bar{D}^2$ interconstellations > $\bar{D}^2$ intra-constellation n'est plus remplie.

façons de l'accroître. Ces $n-2$ accroissements possibles sont à calculer comme suit:

$$\text{Un accroissement} = \left\{ \begin{array}{l} 1^{re} \text{ distance, distance d'un troisième groupe} \\ \quad \text{au premier} \\ + \\ 2^{e} \text{ distance, distance de ce troisième groupe} \\ \quad \text{au deuxième} \end{array} \right.$$

L'accroissement de C1 provoqué par G1 est: distances G1G10 + G1G11=4,91+3,62=8,53; pour G2 c'est G2G10+G2G11=4,07+3,19; etc. Dans cet exemple c'est le groupe 2 qui entraîne l'accroissement minimal 7,26 (le plus élevé des 9 étant 29,89).

C'est le groupe qui provoque l'accroissement minimal qui est retenu, comme troisième groupe de la constellation, *si* toute distance mesurée entre un de ces trois groupes et un des $n-3$ autres groupes dépasse la distance moyenne intra-constellation:

$$\boxed{D^2 \text{ intra-extra-constellation} > \bar{D}^2 \text{ intra-constellation}}$$

Quand, au contraire, une des distances intra-extra-constellation s'avère inférieure à la distance moyenne intra-constellation $\Sigma D^2/3$, il faut en conclure que le troisième groupe est à l'extérieur de la constellation et qu'il n'y a pas d'autre groupe dans la première constellation: les $n-2$ groupes entrent dans d'autres constellations ou restent dissociés.

Le choix, dans l'exemple numérique, de G2 comme troisième groupe de la première constellation conduit au calcul de la distance moyenne intra-constellation 1,28 + 7,26 = 8,54/3 = 2,85. Le diviseur est le nombre de D^2 et non pas le nombre de groupes: trois D^2 (G10G11, G2G10, G2G11) ont donné la somme 8,54. On vérifie si la condition D^2 intra-extra- $> \bar{D}^2$ intra-constellation, imposée à l'adoption de G2, est remplie en parcourant les colonnes G2, G11 et G10 du tableau 1: les D^2 reliant ces groupes aux groupes extérieurs G1, G3 à G9 doivent dépasser 2,85. Ceci est bien le cas pour G2G1 = 7,59, pour G11G1 = 3,62, pour G10G1 = 4,91 et, a fortiori, pour les autres D^2.

Dans l'éventualité de l'adjonction d'un troisième groupe à la première constellation, il y a lieu de calculer, dans une troisième étape, les accroissements de D^2 provoqués par chacun des $n-3$ groupes restants.

$$\text{Un accroissement} = \begin{cases} 1^{re} \text{ distance, distance d'un quatrième groupe au troisième} \\ + \\ 2^{e} \text{ distance, distance de ce quatrième groupe au premier} \\ + \\ 3^{e} \text{ distance, distance de ce quatrième groupe au deuxième} \end{cases}$$

La somme de la deuxième et troisième distances est déjà calculée à l'étape précédente. Le quatrième groupe envisagé pour la constellation est celui entraînant l'accroissement le plus faible. Il n'est retenu que *si toutes les distances intra-extra-constellation sont supérieures* à la nouvelle distance moyenne intra-constellation $\Sigma D^2/6$.

Ceci n'est pas le cas du groupe 1 (cf. tableau 2) de l'exemple, car s'il provoque l'accroissement le plus faible 8,53 + 7,59 (G1G2) = 16,12 des huit accroissements 16,12 28,79 ... 51,49 23,59, la moyenne des six $D^2 = (8{,}54 + 16{,}12)/6 = 4{,}11$ ne remplit pas la condition D^2 intra-extra-constellation $> \bar{D}^2$ intra-constellation; en effet, $G1G3 = 1{,}45 < 4{,}11$. Le dernier groupe G1 est à l'extérieur de la première constellation et la recherche de la première constellation est achevée.

En résumé, le processus d'adjonction de groupes à l'ébauche de constellation imposée par le D^2 minimal est poursuivi jusqu'à l'obtention d'un $\bar{D}^2$ intra-constellation supérieur à un quelconque D^2 intra-extra-constellation. Le dernier groupe envisagé est alors rejeté et le processus recommence à partir d'une ébauche de deuxième constellation.

La deuxième constellation

Le D^2 le plus petit entre les groupes qui n'entrent pas dans la première constellation définit l'ébauche d'une deuxième constellation *si* la distance moyenne interconstellations est supérieure à chacune des deux distances moyennes intra-constellation

$$\boxed{\bar{D}^2 \text{ interconstellations} > \bar{D}^2 \text{ intra-constellation}}$$

Une distance moyenne entre deux constellations s'obtient en divisant la somme des distances, mesurées entre chaque groupe de la première constellation et chaque groupe de la seconde, par le nombre de distances.

L'adjonction éventuelle d'autres groupes aux deux premiers de cette deuxième constellation doit remplir les deux conditions:

$$\bar{D}^2 \text{ interconstellations} > \bar{D}^2 \text{ intra-constellation}$$
$$D^2 \text{ intra-extra-constellation} > \bar{D}^2 \text{ intra-constellation}$$

L'adjonction de groupes est interrompue dès que l'une ou l'autre de ces conditions n'est plus remplie.

G1, G3, G4 à G9 n'entrent pas dans la première constellation (tableau 2). Des colonnes correspondantes du tableau 1, on tire le D^2 minimal = 1,29 = G4G5 qui ébauche une deuxième constellation C2 car la distance moyenne interconstellations C2C1 = (G4G10 + G4G11 + G4G2 + G5G10 + G5G11 + G5G2)/6 = 74,11/6 = 12,35 dépasse les distances moyennes intra-constellation 1,29 (pour C2) et 2,85 (pour C1). Les six accroissements 9,51 à 45,80 désignent G1 comme troisième groupe, mais G1 ne remplit pas la condition D^2 intra-extra-constellation > $\bar{D}^2$ intra-constellation; en effet, G1G3=1,45 < (1,29+9,51)/3= 3,60 et G1 reste à l'extérieur de la deuxième constellation.

Les constellations suivantes

Le D^2 le plus faible entre les groupes se trouvant à l'extérieur des deux premières constellations sert, à nouveau, de point de départ et les deux conditions précitées restent d'application jusqu'à épuisement de tous ces groupes. Les distances moyennes interconstellations sont calculées, par conséquent, lors de chaque tentative d'adjonction de nouveaux groupes.

Six groupes, G1, G3, G6 à G9, sont à l'extérieur de C1, C2. Des six D^2 lus en tête de ces colonnes (tableau 1), 1,45 relie les groupes 1 et 3, ébauche de la première constellation, puisque la condition imposée est remplie:

$$\text{C3C1} = (\text{G1G10} + \text{G1G11} + \text{G1G2} + \text{G3G10} + \text{G3G11} + \text{G3G2})/6 = 44{,}91/6 = 7{,}49 > 1{,}45$$
$$> 1{,}29$$
$$> 2{,}85$$

de même C3C2 > $\bar{D}^2$ intra-constellation

et C2C1 > $\bar{D}^2$ intra-constellation

Quatre groupes subsistent: G6, G7, G8 et G9. Les accroissements de C3 qui résulteraient de leur incorporation à cette constellation sont respectivement

$$\begin{aligned}
&G6G1 + G6G3 = 3{,}36 + 7{,}13 = 10{,}49 = \text{accroissement minimal}\\
&G7G1 + G7G3 = 10{,}25 + 11{,}58 = 21{,}83\\
&G8G1 + G8G3 = 22{,}84 + 29{,}31 = 52{,}15\\
&G9G1 + G9G3 = 16{,}29 + 22{,}34 = 38{,}63
\end{aligned}$$

On vérifie maintenant si G6 entre dans C3 dont la distance moyenne est désormais

$$(1{,}45 + 10{,}49)/3 = 3{,}98$$

Toutes les distances intra-extra-C3 (G6G7 = 6,61, G3G7 = 11,58, G1G7 = 10,25) sont supérieures à 3,98 et toutes les distances moyennes interconstellations (C3C1 = 7,10, C3C2 = 5,87, C2C1 = 12,35) dépassent les distances moyennes à l'intérieur de ces constellations: 3,98; 2,85; 1,29 (cf. tableau 2). G6 est bien dans la constellation C3. Des trois groupes restants, G7 satisfait les mêmes conditions et est retenu. Par contre, de G8 et G9, l'accroissement minimal est provoqué par G9, mais, tout en remplissant la première condition, ce groupe ne remplit pas la seconde et reste à l'extérieur de C3. A la question de savoir si G8 et G9 ne forment pas une quatrième constellation, il est aisé de répondre par la négative, puisque la distance moyenne C4C1 est inférieure à 14,93. Les deux groupes ne présentent aucune affinité.

Le critère de l'étendue, auquel il n'est fait appel qu'une fois, lors de l'acceptation des deux premiers groupes de la première constellation, est arbitraire. Son abandon reviendrait, toutefois, à déceler toujours au moins une constellation, ce qui irait à l'encontre de l'objectif fixé à certaines études, à savoir la recherche du pouvoir maximal de dissociation.

On peut présenter, sous forme matricielle, les distances moyennes intra- en interconstellations (tableau 3). Les distances moyennes interconstellations sont également calculées à partir des accroissements de D^2 du tableau 2. Ainsi,

$$\bar{D}^2 C3C2 = \frac{1}{mn} \sum_{1}^{m} \Delta_{C2} D^2 = (9{,}51 + 12{,}83 + 12{,}90 + 24{,}06)/8 = 7{,}41$$

m et n étant, respectivement, le nombre de groupes de C3 et de C2.

L'interprétation de ces résultats doit porter, à la fois, sur le pouvoir de dissociation des variables, mesuré par les distances respectives entre constellations, entre constellations et groupes individualisés et, aussi, entre groupes individualisés. Elle doit toutefois s'étendre à de grandes séries de tableaux similaires, si l'on veut qu'elle contribue à préciser la classification générale des sols intertropicaux.

Tableau 3. — Distances moyennes intra- et interconstellations $\bar{D}^2$

Constellations	Groupes	C1 G2 +G10 +G11	C2 G4 +G5	C3 G1 +G3 +G6 +G7	G8	G9
C1	G2+G10+G11	2,85	12,35	8,02	17,16	7,86
C2	G4+G5	12,35	1,29	7,41	29,20	22,90
C3	G1+G3+G6+G7	8,02	7,41	6,73	20,90	17,40
	G8	17,16	29,20	20,90	0	14,93
	G9	7,86	22,90	17,40	14,93	0

SUMMARY

Once the (n—1)n/2 *generalized distances* D^2 *of Mahalanobis (1925, 1927, 1930, 1936, 1937, 1948) have been obtained between* n *groups, on* v *variates, there remains the determination of group constellations (Rao, 1952, pp. 361-363).*

A straightforward approach to the clustering process has been followed, enforcing Rao's remark: "The only criterion appears to be that any two groups belonging to the same cluster should at least on the average show a smaller D^2 *than those belonging to two different clusters." The approach may be summarized as follows: To the groups united by the smallest* D^2 *one adds the group giving the smallest increase in* D^2*. The three groups are members of a first cluster if the intra-cluster average distance is smaller than all distances from any of his groups to the remaining groups. Increase in* D^2 *is then repeatedly computed for each of the remaining groups; the smallest increment corresponds to a group which, under the same condition, may be added to the first cluster. The first cluster is augmented likewise until the intra-cluster average* D^2 *becomes larger than any individual* D^2 *from one of his own groups to an extra group. Neglecting the distances measured from the first cluster, the smallest of the remaining distances defines two groups as a second cluster if the inter-clusters average distance exceeds the intra-cluster average* D^2 *values. The group which produces the smallest increment to the second cluster average* D^2 *is accepted, not only under the condition of larger inter-clusters average distance, but also, for the groups not yet clustered, under the above-mentioned condition of larger individual intra-extra-cluster distance. The clustering process is repeated on the remaining groups, bearing in mind the conditions:*

$$\bar{D}^2 \textit{ inter-clusters} > \bar{D}^2 \textit{ intra-cluster}$$

$$D^2 \textit{ intra-extra-cluster} > \bar{D}^2 \textit{ intra-cluster}$$

BIBLIOGRAPHIE

Mahalanobis P.C., "Analysis of race-mixture in Bengal (Presidential address, Anthropology Section, Indian Science Congress, Benares, 1925)", *J. Asiatic Soc. Bengal*, 23, 301-333, 1927.

Mahalanobis P.C., "On tests and measures of group divergence. Part I: Theoretical formulae", *J. Asiatic Soc. Bengal*, 26, 541-588, 1930.

Mahalanobis P.C., "On the generalized distance in statistics", *Proc. Nat. Inst. of Sciences of India*, 2, 49-55, 1936.

Mahalanobis P.C., "Normalization of statistical variates and the use of rectangular co-ordinates in sampling distributions", *Sankhyā*, 3, 35-40, 1937.

Mahalanobis P.C., "Historical note on the D^2-statistic", *Sankhyā*, 9, 237-240, 1948.

Rao C.R., *Advanced statistical methods in biometric research*, New York, Londres, Wiley, 1952, 390 p.