

JACQUES LEFEBVRE

Aspects génétiques des relations entre critères

Journal de la société statistique de Paris, tome 132, n° 2 (1991),
p. 71-78

http://www.numdam.org/item?id=JSFS_1991__132_2_71_0

© Société de statistique de Paris, 1991, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ASPECTS GÉNÉTIQUES DES RELATIONS ENTRE CRITÈRES

JACQUES LEFEBVRE

RÉSUMÉ

Des analyses multidimensionnelles ont permis de mettre en évidence chez des animaux aussi différents que des bovins et des crustacés de quelques millimètres un déterminisme génétique des relations entre critères beaucoup plus précis que celui de ces critères considérés individuellement.

SUMMARY

In animals as different as can be bovine or crustaceans a few millimeters large, multivariate analysis has revealed a genetic determinism for relationships between quantitative traits of a much greater accuracy than achieved when separately considering those traits.

I. INTRODUCTION

Les analyses multidimensionnelles sont souvent utilisées en tant que techniques statistiques descriptives (Thélot C., 1986) faisant ressortir, selon Benzecri J.P. (1973), « les grands courants dans l'océan des faits ».

Elles peuvent cependant aussi être utilisées pour déceler l'existence de relations étroites entre plusieurs critères et pour séparer avec précision des populations selon les niveaux de ces relations.

De tels objectifs ont d'ailleurs été à l'origine de certaines des plus anciennes de ces analyses. C'est ainsi que Mahalanobis proposait le D^2 dès 1927 pour étudier les différences entre races du Bengale et que la fonction discriminante a été conçue par Fisher en 1936 pour des études de taxonomie. Teissier (1938) a utilisé, pour l'étude de la croissance relative chez les crustacés, l'analyse factorielle développée par les psychologues.

Les travaux de Rao (1952) avaient bien mis en évidence l'intérêt des analyses multidimensionnelles en biologie. Mais jusqu'au développement des ordinateurs (vers 1960) leur mise en œuvre se heurtait à la lourdeur des calculs sur machine de bureau. Très utilisées depuis, en particulier pour les études de croissance et de conformation, elles sont par contre pratiquement ignorées

en génétique quantitative, sans doute parce que, depuis Falconer (1960), celle-ci a donné lieu à de nombreux et brillants développements théoriques. Cependant ses concepts de base font l'objet de nombreuses critiques : les effets de seuil contredisent l'additivité et la notion d'héritabilité se révèle très ambiguë (Jacquard A., 1977, 1983). D'autre part, comme l'a souligné Tomassone (1963), il est bien utopique de définir un être vivant par un seul caractère. Les sélectionneurs qui ont créé en un siècle et demi les races modernes ont toujours apporté la plus grande attention aux équilibres entre caractères.

Depuis la dernière guerre les directives officielles prônant toutefois la sélection sur un seul critère (la quantité de lait en particulier chez les bovins) une collaboration étroite a été entreprise en 1958 entre le Professeur C. Bocquet de la faculté des sciences de Caen, la Société de Contrôle des Taureaux Normands (SCTN) et la Station de Recherche sur l'Élevage de l'INRA pour rendre plus efficace la sélection en race Normande. Elle s'est développée ensuite pendant plus de 20 ans avec le Laboratoire de Biométrie et de Génétique Evolutive du CNRS où C. Bocquet avait succédé à G. Teissier, la SCTN et le Laboratoire de Génétique Factorielle du CNRS.

MATÉRIEL ET MÉTHODES

A partir de 1952 les taureaux normands et montbéliards ont fait l'objet de mise à l'épreuve sur descendance dans des conditions particulièrement rigoureuses. C'est ainsi que pendant plus de 10 ans une seule personne, dans chaque race, a effectué les mensurations de 50 à 100 produits par reproducteur (300 normands, 50 montbéliards). Une grande homogénéité dans les données, malgré la grande dispersion géographique des élevages, a pu être ainsi obtenue. Les résultats de production laitière ont également été minutieusement contrôlés par les mêmes personnes. Après 1962 le développement du testage, portant en fin de période d'étude à 2 000 le nombre de reproducteurs et 200 000 celui des animaux examinés n'a pas permis de maintenir cette unicité. Elle a été remplacée par de multiples contrôles de vraisemblances, en particulier à l'aide d'analyses multidimensionnelles préalables. Ces contrôles ont absorbé environ 80% du temps d'ordinateur.

Dans la même période 1952-1962 une centaine de couples de bovins supposés monozygotes ont été suivis pendant toute leur croissance au CNRS, les données étant recueillies également par une seule personne.

Le quatrième fichier utilisé dans cette étude a aussi été constitué au début des années 50. Il comporte 17 mesures corporelles de 1300 Jaera albifrons, crustacé isopode de 2 à 4 mm, c'est-à-dire mille fois plus petits que les bovins. Toutes ont été réalisées par C. Bocquet, essentiellement pour des études d'allométrie (1953). Ces données ont été reprises à la demande de leur auteur en analyses multidimensionnelles. Elles ont également été l'objet de contrôles préalables systématiques par une méthode originale (Boitard M., 1978) qui a permis d'éliminer quelques anomalies généralement dues à des déformations

lors des examens. Dans un article récent, Duprat H. (1990) a bien rappelé l'importance de la précision des mesures.

Les analyses multidimensionnelles sont extrêmement sensibles aux données anormales qui résultent soit d'erreurs soit de causes d'origines diverses-y compris génétiques. Les premières doivent évidemment être éliminées ou corrigées. Quant aux secondes il faut bien les avoir en mémoire lors des interprétations.

Trois analyses ont été utilisées. Celle en composantes principales (Hotelling 1933) est la plus classique et la plus simple malgré bien des controverses au sujet des normes des variables. Son efficacité dans les études d'allométrie généralisée a été peu exploitée.

La fonction discriminante, proposée par Fisher en 1936 comme application de la régression multiple, se prête particulièrement bien à la recherche des variables les plus efficaces. Elle constitue un cas particulier de l'analyse discriminante.

Le D^2 de Mahalanobis, développé par Rao (1952) conduit aux mêmes résultats que cette dernière. Mais dans une première étape, sans calcul de valeurs propres, elle donne les distances entre populations et les tests correspondants, ce qui permet de réunir éventuellement celles qui ne sont pas significativement différentes. La recherche des variables les plus discriminantes y est simple et efficace. Le calcul des vecteurs canoniques de projection, dans une 2^{ème} étape, évite les difficultés de précision numérique rencontrées dans l'analyse discriminante. Il est surprenant que malgré ses avantages la méthode du D^2 soit pratiquement absente des bibliothèques classiques de programmes.

CROISSANCE ET CONFORMATION

Les examens systématiques des descendants des taureaux normands testés à partir de 1952 avaient fait apparaître qu'après avoir vu une quinzaine de produits d'un reproducteur des experts entraînés pouvaient reconnaître les suivants malgré les différences entre les élevages où ils étaient dispersés. Pour chacune des 14 mensurations les moyennes ne présentaient cependant que peu de différences significatives. Les descendances se distinguaient au niveau des interrelations entre l'ensemble des dimensions. Pour les mettre en évidence C. Bocquet a conseillé en 1962, à la suite des travaux de Rao, d'utiliser le D^2 de Mahalanobis. L'installation d'un ordinateur de type Gamma de Bull au Centre de Versailles de l'INRA permettait alors de mettre en œuvre une telle méthode. Entre les descendances de 300 taureaux ayant au total près de 20 000 produits ont pu être mis en évidence de très nombreuses distances très significatives. Les projections sur les 2 premiers axes canoniques, rendant compte des 2/3 de la variabilité, permettaient de définir des zones de conformation caractéristique, traduction des différences génétiques entre reproducteurs (Lefebvre, 1966).

A la suite de ces résultats C. Bocquet demanda que soient reprises ses données concernant 3 espèces de Jaera et 2 de leurs hybrides. Malgré des

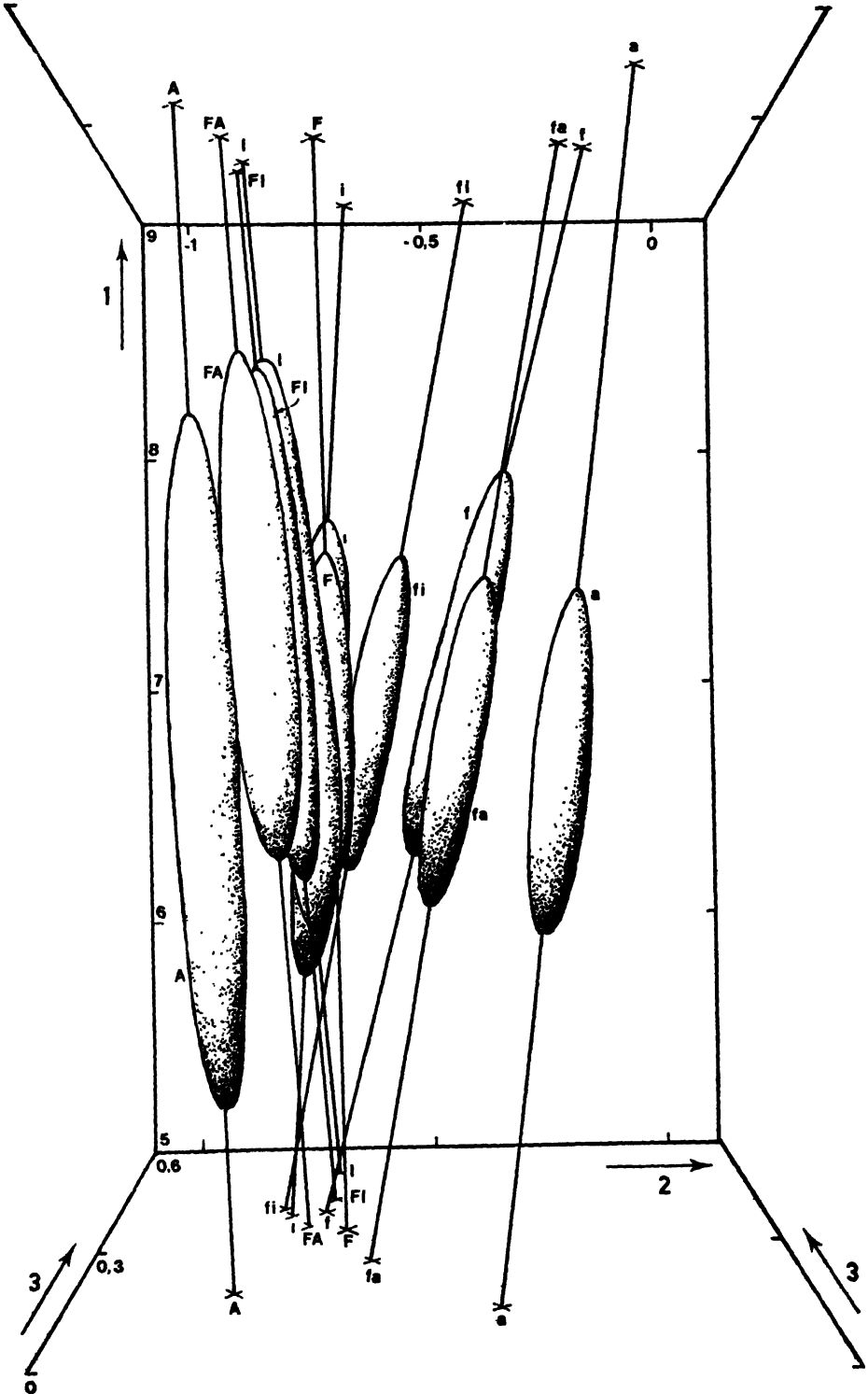
différences considérables de nature et de taille entre ce matériel et le précédent l'analyse a mis également en évidence des différences très significatives de conformation entre populations, alors que la détermination des espèces ne porte que sur les périopodes. Bien qu'elle ne soit possible que chez les mâles la discrimination par le D^2 est presque aussi bonne chez les femelles. Sur le plan des deux premiers axes canoniques les mâles d'une part, les femelles d'autre part sont alignés, l'ensemble formant un V près de la pointe duquel se trouvent les 2 sexes d'une des espèces. Le 1^{er} axe sépare les sexes. Sur le second les deux sexes se projettent dans le même ordre (Boitard M., 1978). Grâce au développement des moyens de calculs le programme de D^2 avait été complété par la recherche systématique des variables les plus efficaces. Six sont apparues, compte tenu des degrés de liberté, aussi discriminantes que les 17 d'origine. Elle correspondent donc aux critères dont les relations sont le plus étroitement régies par la constitution génétique propre à chaque espèce.

L'analyse discriminante annulant pratiquement les différences de taille, l'étude des mêmes populations a été reprise en composantes principales, ce qui constitue une généralisation des relations d'allométrie étudiées par C. Bocquet en 1953. La première composante, facteur générale de taille, est 10 à 30 fois plus grande que les 2^{ème} et 3^{ème}. Toutes les variables d'origine ont une corrélation supérieure à 0,99 avec elle : pendant la phase de croissance, comportant un grand nombre de stades, les dimensions s'accroissent avec une remarquable uniformité malgré les changements de proportions résultant des allométries. Les 1 300 individus ayant été projetés dans l'espace de ces 3 premières composantes, on vérifie tout d'abord que les dispersions de chaque population sont très proches des ellipsoïdes théoriques. Ceux-ci sont tous distincts, plus chez les mâles que chez les femelles, où, nous l'avons signalé, l'espèce ne peut être déterminée. Au cours de la croissance les directions générales de croissance des populations s'écartent, leurs angles allant de 0,5 à 10,7 degrés. Ces premières composantes sont sensiblement dans deux plans, un pour chaque sexe. La 2^{ème} composante sépare les sexes. Sur la 3^{ème} mâles et femelles d'une même population se projettent sensiblement au même point. La figure 1 restitue de façon frappante la rigueur avec laquelle la conformation propre à chaque espèce et chaque hybride est maintenue tout au long de la croissance.

L'efficacité de l'allométrie généralisée s'est retrouvée dans l'étude biométrique des jumeaux monozygotes. Sur les 83 paires qui ont pu être suivies sur une assez longue période 10 se sont finalement révélées fausses et 41 douteuses selon divers tests. Une analyse en composantes principales a été effectuée sur la première période (de la naissance à la première rupture de pente, soit environ vers 14 mois), un test précoce de monozygotie étant important. Entre les 1^{ères} composantes des deux éléments de chaque couple l'angle varie de 3° à 10° pour les monozygotes et de 15° à 30° pour les dizygotes (Lefebvre *et al.*, 1974).

FIGURE 1

Ellipsoïdes de dispersion des projections,
dans l'espace des 3 premières composantes principales,
*des 3 espèces (1 lettre) et des 2 hybrides (2 lettres) de *Jaera marina*.*
Les mâles sont en minuscules, les femelles en majuscules.
L'échelle de la 1^{ère} composante a été divisée par 2.



DISTANCES ET PARENTÉS

Le problème des liaisons éventuelles, favorables ou défavorables, entre conformation et production laitière a été abordé avec des mesures nombreuses et précises il y a plus de 150 ans. Depuis beaucoup de corrélations ont été calculées entre critères des deux catégories. Les valeurs trouvées ont été très diverses et ont rarement dépassé 0,2 ou 0,3. Pour 14 à 17 mensurations et une dizaine d'éléments de production laitière chez les produits de 40 taureaux montbéliards et 44 normands, les corrélations canoniques ont été dans les deux cas sensiblement égales à 0,8, valeur très supérieure au seuil $P = 0,01$. Cette coïncidence est remarquable pour des races d'origines et de conditions de vie très différentes. Pour la normande chaque descendance a été projetée sur le plan des premières variables synthétiques lait et conformation. Sont alors apparues très proches celles des reproducteurs très apparentés (frères ou 1/2 frères, père-fils...). Ici l'action du génotype commun se manifeste sur les liaisons entre deux catégories très différentes de critères (Lefebvre J., 1967).

Les problèmes de précisions évoqués précédemment se sont manifestés de façon particulièrement nette au cours de cette étude. Dans les calculs numériques tout d'abord, les programmes alors disponibles étant loin de fournir entre les projections des corrélations réelles égales aux corrélations canoniques théoriques (Lefebvre J., Tassencourt L., 1968). Pour la prise de données d'autre part : lorsque le testage s'est développé les animaux n'ont pu être examinés par la même personne et en recalculant les corrélations canoniques on a constaté que leurs valeurs diminuaient légèrement et progressivement, tout en restant très significatives. Ceci a été logiquement attribué aux différences entre examinateurs.

La confirmation d'une relation entre distance et parenté a été recherchée ultérieurement par une autre voie. En 1980, sur 1660 taureaux connus, 28 avaient eu de 8 à 71 fils eux-mêmes testés (au total 669). Les 378 parentés entre ces 28 têtes de lignées ont été calculées sur 6 générations. D'autre part un D^2 progressif entre les 28 groupes de fils, chacun étant défini par la moyenne de ses filles (de 28 à 212, soit 53 682 au total), a conduit à retenir les 7 variables les plus efficaces. Aucun index ne figure parmi elles alors qu'ils sont théoriquement la meilleure expression du génotype. Une représentation des 378 D^2 en fonction des parentés a fait apparaître une nette diminution des distances pour des parentés supérieures à 2%. Les distributions des D^2 pour les deux classes de parenté montrent des différences très significatives de moyennes et de dispersions (Tableau I). Le programme de parenté donnant la contribution de chaque ancêtre, dans 78 des 105 supérieures à 2% deux reproducteurs sont apparus intervenir de façon prépondérante. Pour ce sous-groupe la différence avec les 273 est encore plus grande et plus significative. Ainsi a été retrouvée la notion de « raceur » familière aux sélectionneurs (Lefebvre *et al.*, 1982). Une étude analogue réalisée avec les mensurations a donné des résultats équivalents mais la différence entre les groupes n'a pas été supérieure alors que sa conformation est classiquement considérée comme plus « héritable » que la production laitière.

TABLEAU I
*Test des différences entre les moyennes
 et les dispersions des groupes de D^2 , selon les parentés*

Parentés	$p > 2\%$	
	$n = 105$	Sous-groupe $n = 78$
$p \leq 2\%$ $n = 273$	$t = 4,16^{++}$ $F = 1,46^{++}$	$t = 6,54^{++}$ $F = 2,81^{++}$

FACTEURS MENDÉLIENS ET PRODUCTION LAITIÈRE

En dernier lieu les relations entre critères de production ont pu être étudiées dans la même base de données en fonction de facteurs mendéliens. En effet la détermination des groupes sanguins est obligatoire pour tous les taureaux d'insémination. Les 76 facteurs sanguins des bovins se répartissent en 11 systèmes génétiques qui en comprennent de 1 à 42. Jusqu'à présent aucune liaison certaine n'a pu être établie entre 1 facteur sanguin et 1 critère de production laitière. Aussi l'étude a-t-elle été reprise avec les phénogroupes, ensembles de deux à une quinzaine de facteurs d'un même système, cet ensemble, pour un animal donné, se transmettant en bloc. Les phénogroupes sont en nombres très variables. Il y en a par exemple plus de 250 connus en race normande au système B.

Parmi les têtes de lignées précédemment étudiées, 15 ont été retenues comme ayant plus de 30 fils afin d'avoir des effectifs suffisants pour chaque phénogroupe. Actuellement les résultats les plus intéressants ont été obtenus avec les systèmes B et C pris séparément puis simultanément. Des D^2 progressifs ont donné dans de nombreux cas des distances très significatives avec 2 ou 3 critères de quantité et (ou) de qualité. Ces résultats méritent d'autant plus l'attention que chaque phénogroupe n'a qu'une chance sur deux d'être transmis à chacune des petites filles et les groupes sanguins ne sont malheureusement déterminés que par sondage chez les femelles.

Les corrélations canoniques avaient mis en évidence d'étroites liaisons entre deux combinaisons de caractères quantitatifs. Ici la relation apparaît entre de tels critères et un ensemble de facteurs mendéliens.

CONCLUSION

Ces différents résultats montrent que les variables canoniques ne sont pas de simples abstractions algébriques mais l'expression de facteurs génétiques qui régissent très étroitement les relations entre les critères mesurés. Ces relations pouvant être très différentes selon les populations (espèces, races,

lignées...) il ne faut tirer aucune conclusion définitive de simples calculs de corrélations globales. La sélection sera d'autant plus efficace qu'elle reposera sur les structures révélées par les analyses multidimensionnelles, de même que la taille des pierres ne peut s'opérer qu'en tenant compte des plans de clivage.

BIBLIOGRAPHIE

- BENZECRI J.P. (1973) L'analyse des données, II. L'analyse des correspondances, 619 pp., Dunod, Paris.
- BOCQUET C. (1953) Recherches sur le polymorphisme naturel de *Jaera marina* (Fabr.) (Isopodes Asellotes). Arch. Zool. exp. gén., 90, 187-450.
- BOITARD M. (1978) Application des méthodes d'analyses multidimensionnelles à l'étude du complexe *Jaera albifrons* (Isopodes Asellotes). Thèse Doc. Univ., Paris VI, 116 pp.
- DUPRAT H. (1990) Sciences des mesures et mesure des sciences. Journ. Soc. Stat. Paris, 131, 79-97.
- FALCONER D.S. (1960) Introduction to Quantitative Genetics. London, Oliver and Boyd.
- FISHER R.A. (1936) The use of multiple measurements in taxonomic problems. Ann. Eugen., 7, 179-188.
- JACQUARD A. (1977) La génétique peut-elle être quantitative? La Recherche, 79, 590-591.
- JACQUARD A. (1983) Heritability : One Word, Three Concepts. Biometrics, 39, 465-477.
- HOTELLING H. (1933) Analysis of a complex of statistical variables into principal components. J. Educ. Psychol., 24, 417-441 et 498-520.
- LEFEBVRE J. (1966) Etude, à l'aide de mensurations, de la conformation et de la croissance des bovins normands. Thèse Doc. Ing. Fac. Sci. Uni. Caen, 124 pp.
- LEFEBVRE J. (1967) L'analyse canonique. Séminaire de Biométrie, Nancy, mars.
- LEFEBVRE J., BOITARD M., REY J.-F. (1981) Logiciels d'analyses statistiques multidimensionnelles sur micro-ordinateurs, 152 pp., ed. Masson, Paris.
- LEFEBVRE J., AURIOL P., de PREMÉSNI J., DUPONT M. (1974) Analyse multidimensionnelle de la croissance et de la conformation des bovins jumeaux monozygotes. 1^{er} Congrès Mondial de la Génétique appliquée à l'élevage. Madrid.
- LEFEBVRE J., HAMON M., CALENS V., WIMITZKY M. (1982) Parentés entre 28 taureaux normands têtes de lignées et distances entre leurs 669 fils testés. XI^{ème} Conf. Int. de Biométrie, Toulouse, 6-11 septembre.
- LEFEBVRE J. (1983) Introduction aux Analyses Statistiques Multidimensionnelles, 275 pp, 3^{ème} ed., Masson, Paris.
- LEFEBVRE J., TASSENCOURT L. (1968) Le problème de la précision dans le calcul des corrélations canoniques. Biom. Prax., 9 (3), 161-173.
- MAHALANOBIS P. (1927) Analysis of Mixture in Bengal. J. Asiat. Soc. Bengal, 23, 301-333.
- RAO C.R. (1952) Advanced Statistical Methods in Biometric Research, 390 pp., John Wiley, London, New-York.
- TEISSIER G. (1938) Un essai d'analyse factorielle. Les variants sexuels de *Maia Squinado*. Biotypologie, 6, 73-97.
- THÉLOT Cl. (1986) La statistique, science de la mesure. Journal de la Société de Statistique de Paris, 67-85.
- TOMASSONE R. (1963) Application des fonctions discriminantes à des problèmes biométriques. Ann. Ec. Nat. Eaux et Forêts, 20, 585-617.