

Discriminations et classifications basées sur des caractéristiques du status bucco-dentaire

R. Lang et D. Ramaciotti

Institut de médecine sociale et préventive
1205 GENEVE

1. Introduction : Problématique et méthode

L'analyse discriminante est une technique statistique multivariée : elle présuppose que, sur chaque membre d'une collectivité C sans exception, on ait mesuré la valeur numérique de m variables quantitatives x, y, z, etc.. distinctes ("variables discriminatoires"). Pour pouvoir être appliquée, cette technique nécessite en outre une partition du collectif C en plusieurs groupes mutuellement exclusifs C', C'', C''', etc., définis a priori par des critères indépendants des variables discriminatoires. Dans ces circonstances, l'analyse discriminante tente de répondre à des questions du genre suivant:

- Quelle est la combinaison linéaire ("fonction discriminante") $f(x; y; z; \dots) = ax + by + cz + \dots$ qui discerne de façon optimale les groupes C', C'', C''', etc ? Autrement dit, quels sont les facteurs de pondération a, b, c, etc., pour lesquels les scores individuels w_i

$$w_i = ax_i + by_i + cz_i + \dots \text{ (membre no } i \text{)}$$

présentent une faible dispersion intra-groupes mais une forte dispersion inter-groupes ?

- Quelles sont, parmi les variables discriminatoires x, y, z, etc., celles qui contribuent le plus efficacement à cette dissociation ? Quelles sont celles qui pourraient, le cas échéant, être omises ?

- Un groupe indéterminé C* comprenant tous les membres pour lesquels l'information classificatrice -celle permettant de les classer a priori dans l'un des groupes C', C'', etc.- manque, peut-on reclasser a posteriori les membres de C* dans l'un approprié des groupes C', C'', etc., et ce avec quel risque d'erreur ?

Par rapport à une succession de m analyses univariées indépendantes pour chacune des variables discriminatoires séparément, cette technique de discrimination présente l'avantage d'incorporer dans une unique analyse ou un unique procédé classificateur l'ensemble des variables discriminatoires, compte tenu de leurs liaisons (corrélations) et de leur pouvoir discriminatoire.

Pour le détail méthodologique, cf. (1), (2)

2. Spécification des données : groupes et variables

Les données numériques analysées ici émanent toutes de l'"Enquête sur la denture et le parodonte des adolescents de 16 à 20 ans domiciliés à Genève (1970-1972)", basées sur un échantillon aléatoire de garçons et de filles, sans distinction d'origine, de nationalité, de milieu socio-culturel ou de niveau économique ; effectif de l'échantillon : N = 1392, soit 13 % de la population adolescente du canton de Genève (3). Sans égard à l'âge ni au sexe, ce collectif peut être fractionné selon l'occupation (étudiants : 53,9% de N; apprentis : 26,9%; jeunes travailleurs : 19,2%) et, indépendamment de celle-ci, selon la durée d'absorption de comprimés fluorés à titre prophylactique (aucune absorption : 18,2% de N; absorption pendant moins de 5 ans : 40,1%; pendant plus de 5 ans : 25,4%; groupe indéterminé : 16,2%).

Les 11 variables discriminatoires intervenant ici spécifient (I) l'atteinte carieuse (nombre de surfaces dentaires saines, de surfaces cariées de gravité 1, 2, ou 3, de surfaces obturées, de surfaces manquantes/à extraire pour cause de carie) ; (II) l'hygiène bucco-dentaire (tartre; débris/taches) ; (III) l'état du parodonte (aux papilles gingivales ; aux bords marginaux de la

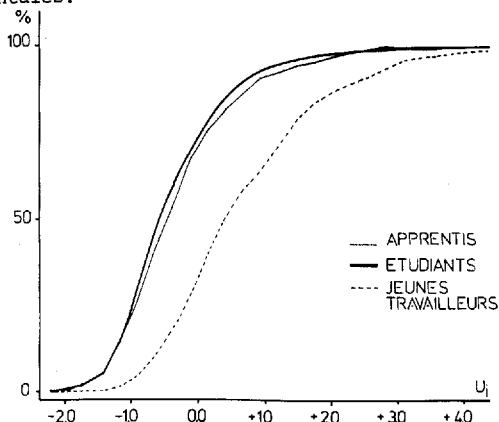
gencive) ; enfin (IV) la présence de fractures (nombre de dents antérieures fracturées).

3. Discrimination des occupations

Dans cette discrimination, on a attribué aux sous-échantillons "Étudiants" et "Jeunes travailleurs" le rôle des groupes définis a priori C', C'' respectivement, tandis que le sous-collectif "Apprentis", bien que parfaitement déterminé lui aussi, fonctionnait provisoirement comme groupe indéterminé C* à reclasser a posteriori. Cette approche était motivée par l'incertitude quant à la justification d'une partition en 3 groupes occupationnels. Souvent, en effet, un apprenti présente les mêmes particularités culturelles, socio-économiques, d'origine ou d'instruction qu'un étudiant ; dans le présent contexte, la distinction entre étudiant et apprenti pourrait, de ce fait, paraître artificielle.

Pour toutes les variables mentionnées sauf les fractures, le test univarié de comparaison des moyennes indique un écart significatif (seuil de probabilité 1%) entre étudiants et jeunes travailleurs. Rangées selon l'importance décroissante de ces écarts, les variables se suivent dans l'ordre tartre, surfaces manquantes/à extraire, surfaces cariées de gravité 2, saines, cariées de gravité 3 et de gravité 1, surfaces obturées, débris/taches, bords marginaux, papilles gingivales, dents fracturées.

Ces 11 variables permettent conjointement d'établir une discrimination statistiquement significative (seuil 1%) entre étudiants et jeunes travailleurs. Cependant, 5 seulement parmi elles ont un pouvoir discriminatoire significatif (seuil 1%) ; ce sont, par ordre décroissant : tartre, surfaces manquantes/à extraire, surfaces cariées de gravité 2, surfaces obturées et surfaces cariées de gravité 1. Les 3 premières sont aussi celles qui, dans la fonction discriminante f, sont pondérées le plus lourdement ; viennent ensuite, dans f, les deux variables parodontales.



La figure ci-dessus montre la répartition des scores individuels w_i donnée par la fonction f complète (11 variables) ; il s'agit des fréquences cumulées relatives, en % de N. Les apprentis sont très proches des étudiants ; en effet, selon cette discrimination, 92,3% des apprentis seraient reclassés parmi les étudiants, contre 7,7% affectés aux jeunes travailleurs. Pour ce reclassement, on a observé le schéma décisionnel suivant : si le score individuel w_i d'un membre indéterminé est inférieur (respectivement supérieur) à w^* , ce membre est affecté aux étudiants (respectivement aux jeunes travail-

leurs), w^* étant un "score critique" calculé en cours d'analyse sur la base de considérations probabilistes. Dans cette discrimination, le risque estimé de classification erronée s'élève à 22,0 %.

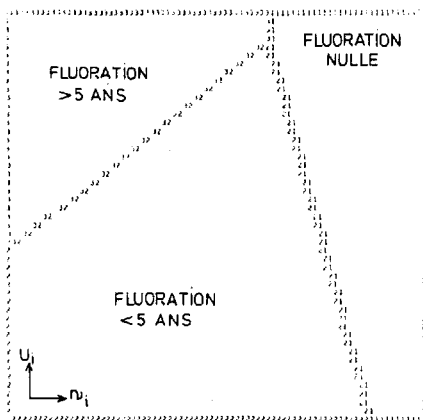
4. Discrimination relative au fluor

Dans ce cas, ce sont les 3 sous-échantillons déterminés par la durée de la prophylaxie par fluoruration qui tiennent lieu de groupes déterminés a priori : aucune absorption (C'), prophylaxie pendant > de 5 ans (C''), pendant < de 5 ans (C'''). Le groupe à proprement parler indéterminé C* est constitué par 226 adolescents (16,2% de N) examinés mais qui n'ont pas pu fournir d'information quant à la prophylaxie en question. Inclassables a priori, ces adolescents étaient à reclasser a posteriori pour pouvoir être pris en considération à part entière dans d'autres investigations.

Les 11 analyses de variance univariées simples indiquent des écarts statistiquement significatifs (seuil 1%) entre les moyennes des 3 groupes C', C'', C''' pour les variables suivantes, par ordre d'importance décroissante : surfaces manquantes/à extraire, tartre, surfaces cariées de gravité 2, celles de gravité 1 et surfaces saines.

Ces 11 variables discriminatoires permettent conjointement d'établir une discrimination statistiquement significative (seuil 1%) entre les 3 groupes de fluoruration déterminés. En présence de 3 groupes déterminés, on obtient plus précisément - de par la nature de la mathématique sous-jacente - 2 fonctions discriminantes distinctes capables chacune de discriminer significativement les 3 groupes de prophylaxie. Les variables discriminatoires contribuant significativement (seuil 1%) à la séparation de ces 3 groupes sont, par ordre d'importance décroissante : surfaces manquantes/à extraire, tartre, surfaces cariées de gravité 1 et 2 (viendraient ensuite les surfaces saines, significatives à 5%). Les 4 premières sont aussi celles qui, dans la première fonction discriminante, sont pondérées le plus lourdement; viendraient ensuite, dans cette fonction : papilles gingivales, débris/taches, bords marginaux. Dans la deuxième fonction discriminante, par contre, les variables se succèdent comme suit par ordre de pondération décroissante : variables parodontales, surfaces saines et manquantes/à extraire, tartre, et ainsi de suite.

Les deux fonctions discriminantes attribuent à chaque membre un couple w_i, u_i de scores individuels pouvant être interprétés comme coordonnées d'un point dans un graphique à deux axes. Du point de vue classification



a posteriori, cela revient à délimiter sur ce graphique 3 zones d'affectation correspondant respectivement aux groupes déterminés C', C'', et C''', cf. figure ci-contre. Un membre du groupe indéterminé C* (durée de fluoruration inconnue) serait dès lors affecté au groupe C', C'' ou C''', selon que le point représentatif de son couple de scores w_i, u_i tombe dans la zone relative à C', C'' ou C'''. En l'occurrence, ce procédé reclasse les 226 adolescents indéterminés comme suit : 4,4% dans C' (pas de prophylaxie), 3,5% dans C'' (fluoruration > 5 ans) et 92,0% dans le groupe intermédiaire C''' (fluoruration < 5 ans). Dans cette discrimination, le risque estimé de classification erronée s'élève à 51,4%.

5. Conclusions

Les 11 caractéristiques bucco-dentaires considérées ici dissocient donc, ensemble et significativement, aussi bien les 3 groupes occupationnels (étudiants, apprentis, jeunes travailleurs) que les 4 groupes de prophylaxie par le fluor (fluoruration nulle, pendant moins ou plus de 5 ans, indéterminée). A cet égard, l'hygiène bucco-dentaire ainsi que les caries de gravité 1 ou 2 sont déterminantes. Dans l'interprétation de l'importance que les tests statistiques accordent aux surfaces manquantes ou à extraire, il convient de préciser toutefois que sur l'ensemble des 10344 dents, seules 78 ont été indiquées comme manquantes et 2 à extraire pour cause de carie.

Zusammenfassung

Trennverfahren und Zuordnung aufgrund von zahnmedizinischen Beobachtungen

Zwei praktische Anwendungen des statistischen Trennverfahrens und der damit verbundenen Zuordnungsmethode werden anhand von zahnmedizinischen Beobachtungen (Karies Mundhygiene, Zahnbettbefund, Zahnbruch) dargestellt und die relative Wichtigkeit der einbezogenen Veränderlichen besprochen.

Summary

Discriminant analysis and classification based on dental health data

Discriminant analysis and its corresponding classification procedures are illustrated by means of two practical applications pertaining to a dental health survey. The discriminating variables refer to dental caries, oral hygiene, periodontal diseases and fracture of anterior teeth. Their relative discriminating power is discussed.

Bibliographie

- (1) TATSUOKA, M.M. : Multivariate analysis. John Wiley and Sons. New York (1971).
- (2) NIE, N.H. et al. : Statistical package for the social sciences. McGraw-Hill. New York (1975).
- (3) LANG, R., RAMACIOTTI, D. : Le coût de la restauration fonctionnelle bucco-dentaire chez les adolescents de 16 à 20 ans domiciliés à Genève (Enquête 1970-72). Rev.Epidém.et Santé Pub., 1977,25, 41-65.

Adresse des auteurs

R. Lang, D. Ramaciotti, Institut de médecine sociale et préventive, 20, quai Ernest-Ansermet, CH-1205 GENEVE