Connaissance Vigne Vin, 1987, **21,** N° 1, 43-57.

APPLICATION DE L'ANALYSE MULTIDIMENSIONNELLE DES DONNEES A LA RECONNAISSANCE DES VINS ROUGES DE LA RIOJA

Rosa M. TAPIAS, Pilar CALLAO, Maria S. LARRECHI, J. GUASCH et F. X. RIUS

Unité de Chimie analytique, Faculté de Chimie de Tarragone, Université de Barcelone, 43005 Tarragone (Espagne)

INTRODUCTION

Il est important, en œnologie, de trouver des critères objectifs permettant de reconnaître les vins entre eux. Le but est double : d'une part influer sur la qualité en optimisant l'utilisation des techniques œnologiques ; d'autre part détecter toute contrefaçon d'un produit d'appellation.

Encore récemment seuls les critères de la dégustation étaient pris en compte afin de juger la qualité d'un vin et d'en reconnaître l'origne, voire l'âge. Ce mode de jugement a cependant ses limites en raison de son côté subjectif. L'introduction des méthodes d'analyse des données va dans le sens d'une rationalisation dans le domaine de l'appréciation des vins. Les variables deviennent alors des paramètres physico-chimiques parfaitement connus. Il est alors possible d'établir une corrélation entre ces données et des propriétés difficiles à cerner comme l'origine d'un vin ou sa « qualité ».

Ces méthodes nouvelles ont été déjà appliquées au domaine œnologique : reconnaissance de vins d'origines géographiques distinctes (MEDINA et *al.,* 1984 ; VAN DER VOET et *al.,* 1984 ; MORET et *al.,* 1984 ; FRANK et KOWALSKI, 1984) ; reconnaissance de vins de cépage (RAPP et GUNTERT, 1985 ; MORET et *al.,* 1985 ; POLO et *al.,* 1984). Malgré la persistance d'un certain risque d'erreur, la détection des fraudes au niveau des appellations est grandement facilitée par ces techniques. Il en est de même lorsqu'il s'agit d'établir une relation entre la composition d'un vin et sa qualité (NOBLE et *al.,* 1984 ; SHINOHARA, 1984 ; BERTUCCIIOLI et *al.,* 1984).

Ces méthodes ont été appliquées par TAPIAS (1985) à la différenciation et au classement de quatre types de vins rouges espagnols. Les va-

Les quatre régions espagnoles qui ont été testées au cours de cet essai sont éloignées les unes des autres ; par conséquent, les résultats de différenciation statistique sont à relier aux importantes variations de certains paramètres étudiés, comme les données climatiques, la composition des sols, les cépages concernés, etc.

Dès lors, il est apparu intéressant, d'étudier la valeur de telles variables lorsqu'il s'agit d'effectuer une différenciation de vins provenant d'appellations voisines, donc de caractères assez proches. L'expérience a été menée à partir de vins rouges de la Rioja (millésimes 1976 et 1977), provenant des trois appellations Rioja : Rioja d'Alava, Basse Rioja et Haute Rioja.

MATERIEL ET METHODES

Soixante-seize vins (les « objets ») de la Rioja ont été étudiés. Les caractéristiques physico-chimiques de ces derniers ont été publiés (INIGUEZ, 1980).

La répartition en appellations (les « classes ») est la suivante : Rioja d'Alava, 39 échantillons ; Basse Rioja, 13 échantillons ; Haute Rioja, 24 échantillons. Les millésimes étudiés sont 1976 et 1977.

Les paramètres physico-chimiques pris en compte (les « variables ») sont les suivants : densité (DEN), degré alcoolique (GRA), acidité volatile (AV), acidité totale (AT), extrait sec (EXT), cendres (CEN), alcalinité des cendres (ACEN), matières réductrices (MR), indice de Folin (IF), acide tartrique (ACT), acide lactique (ACL), potassium (K), glycérol (GL), butanediol-2,3 (BUT), proline (PR) et nitrates (N).

Les méthodes d'analyse sont les méthodes officielles correspondantes. Il n'a pas été tenu compte des valeurs des pH ni des teneurs en substances azotées, non connues pour tous les vins étudiés.

Quatre méthodes d'analyse des données ont été utilisées. La première est l'analyse linéaire statistique discriminante (SLDA). Cette méthode (COOMANS et *al.*, 1979) permet de calculer les fonctions discriminantes qui peuvent s'associer à des axes placés dans l'espace n-dimensionnel. Ces fonctions qui séparent les différentes classes sont le résultat de pondérations et de combinaisons linéaires des variables de telles façon que la projection sur ces axes des échantillons qui forment chaque classe (chacun étant considéré comme un point dans l'espace n-dimensionnel) présente une séparation maximale. La mise en œuvre de cette technique a été faite au moyen des sous-programmes CONDESCRIPTIF et DISCRIMINANT du logiciel SPSS (NIE et *al.*, 1975). Les calculs ont été effectués sur un DCT 2 000 au Centre de Calcul de l'Université Polytechnique de Barcelone, relié à un calculateur UNIVAC 1 100/80 du Ministère de l'Education et de la Science de Madrid.

La deuxième méthode des K plus proches voisins (KNN) est basée sur la mesure des distances entre objets analysés. La méthode KNN calcule la matrice de distance qui contient toutes les distances entre les points (les vins) placés dans l'espace n-dimensionnel des variables. Un vin inconnu y est lors inclus dans la classe à laquelle appartient la majorité de ses K voisins les plus proches.

La troisième méthode, la machine d'apprentissage linéaire (LLM) est basée sur le calcul d'une surface de classification binaire (VARMUZA, 1980). Cette surface, placée dans l'espace n-dimensionnel défini par les variables est calculé de façon intérative en utilisant la méthode d'erreur et de correction appliquée aux vins d'origine connue. Dans cette méthode on n'utilise pas les paramètres statistiques des classes considérées. Les objets inconnus sont classés selon leur situation dans un des côtés de l'hyperplan calculé.

Enfin, la quatrième méthode est celle bien connue de l'analyse en composantes principales (ACP) (MASSART, 1983). Cette technique permet de visualiser dans un espace inférieur au égal à trois dimensions la disposition originale des vins placés dans l'espace n-dimensionnel initial. Les nouveaux axes qui forment l'espace réduit sont calculés par combinaison linéaire des variables d'origine, de telle façon que la dispersion maximale des données initiales soit représentée dans le nouvel espace.

Ces trois dernières méthodes utilisent les sous-programmes KNN, PLANE et KAPRIN-KATRAN du logiciel de traitement des données ARTHUR (HARPER et *al.*, 1977). Les calculs ont été effectués sur un ordinateur IBM 3085, au Centre d'Informatique de l'Université de Barcelone.

RESULTATS ET DISCUSSION

Le tableau I rassemble les résultats des analyses de chaque groupe étudié et indique, pour chaque cas, la valeur de la variance.

On observe des différences statistiquement significatives pour toutes les mesures effectuées, excepté pour l'acide lactique (ACL), les nitrates (N) et, dans une moindre mesure, les cendres (CEN), le glycérol et la proline (PR).

Il existe alors une probabilité de 15,8 p. 100 d'obtenir une relation F favorable à l'identification exacte d'un vin, lorsque les moyennes des mesures sur les trois groupes coïncident.

Si l'on compare ces résultas à ceux obtenus par TAPIAS (1985) pour les mêmes variables, on observe une nette diminution des valeurs de la releation F. Ceci s'explique par le fait que les vins analysés sont plus proches les uns des autres, donc en fin de compte plus difficiles à différencier. Cependant, il faut remarquer qu'il existe des différences appré-

TABLEAU I

Composition analytique des vins de l'appellation Rioja

Mesure	Valeurs moyenes à 95 % de confiance				
	Haute Rioja[a]	Basse Rioja[b]	Rioja d'Alava[c]	Relation F[e]	Signification
Densité 20/20 (DEN)	994.5 ± 0.5	991.6 ± 0.6	992.3 ± 0.5	6.11	0.004
Titre alcoolique (GRA) % (V/V)	11.3 ± 0.4	15.2 ± 0.5	12.8 ± 0.3	20.38	0.000
Acidité volatile (AV) (g/l H_2SO_4)	0.36 ± 0.05	0.42 ± 0.06	0.31 ± 0.06	0.69	0.000
Acidité totale (AT) (g/l H_2SO_4)	3.84 ± 0.23	3.44 ± 0.24	2.63 ± 0.18	23.50	0.000
Extrait sec (EXT) (g/l)	25.8 ± 0.7	29.5 ± 0.8	23.7 ± 0.6	11.66	0.000
Cendres (CEN) (g/l)	2.58 ± 0.06	2.37 ± 0.07	2.45 ± 0.05	1.65	0.201
Alcalinité des cendres (meq/l K_2CO_3)	25.7 ± 0.6	21.2 ± 0.5	24.2 ± 0.5	8.36	0.001
Matières réductrices (MR) (g/l)	1.69 ± 0.12	2.31 ± 0.15	1.41 ± 0.13	10.56	0.000
Indice de Folin (IF)	40.0 ± 0.9	52.0 ± 0.8	45.0 ± 0.9	13.14	0.000
Acide tartrique (ACT) (g/l)	1.7 ± 0.1	1.5 ± 0.1	1.2 ± 0.1	10.32	0.000
Acide lactique (ACT) (g/l)	1.7 ± 0.1	1.9 ± 0.1	1.8 ± 0.1	0.57	0.567
Potassium (K) (mg/l)	1057 ± 29	861 ± 26	1083 ± 31	13.59	0.000
Glycérol (GL) (g/l)	7.2 ± 0.21	8.1 ± 0.19	7.5 ± 0.20	1.51	0.228
Butanediol-2,3 (BUT) (g/l)	0.60 ± 0.04	0.75 ± 0.05	0.55 ± 0.04	9.32	0.000
Proline (N) (mg/l)	962 ± 0.1	797 ± 11	1029 ± 0.2	1.90	0.158
Nitrates	1.7 ± 0.1	1.6 ± 0.3	1.6 ± 0.2	0.08	0.919

a : 24 échantillons, récoltes 1976 et 1977; b : 13 échantillons, récolte 1976; c : 39 échantillons, récoltes 1976 et 1977; e : analyse de la variance avec 2 et 75 degrés de liberté.

ciables entre les trois groupes étudiés ici, d'un point de vue du degré alcoolique (GRA), de l'acidité totale (AT), de l'extrait sec (EXT), de l'indice de Folin (IF) et de la teneur en potassium (K). Ces différences ont trois origines principales : le cépage différent d'une région à l'autre, les modes de conduite de la vigne, les variations climatiques. Pour l'heure il n'est pas possible de pondérer ces facteurs de variation dans la valeur finale des résultats.

Le degré alcoolique est une des variables parmi les plus significatives quant à la distinction des trois zones, la Basse Rioja produisant les vins à plus haut degré. En effet, le Grenache y est le cépage majoritaire et sa richesse en sucre est supérieure à celle du Tempranillo, principal cépage des régions de Haute Rioja et Rioja d'Alava.

De la même façon, pour les trois régions, les données climatiques vont dans le même sens que celui indiqué plus haut. La Basse Rioja est soumise à un microclimat de type méditerranéen, assez sec, donc l'effet sur une maturité élevée est incontestable. Les deux autres régions sont plus humides, avec un maximum de précipitations au printemps et à l'automne, conditions qui se rapprochent des caractéristiques d'un climat océanique comme celui du golfe de Gascogne.

Par ailleurs, on remarque que les vins des trois régions ont des niveaux différents en potassium. Ceci s'explique par un milieu plus fortement alcoolisé qui va dans le sens d'une plus grande insolubilité du bitartrate de potassium et donc d'une teneur en potassium plus faible. C'est ce que l'on rencontre dans les vins de la Basse Rioja bien qu'il faille se garder de toute conclusion hâtive, étant donné l'influence, au cours de la vinification, de divers paramètres sur la teneur en cet ion (AMERINE, 1980). A la différence d'autres métaux comme le lithium dont la teneur dans le sol a été corrélée à la teneur dans le vin (BERMEJO, 1974), cette relation est très difficile à établir pour le potassium. Les différence des teneurs en matières réductrices varient dans le même sens que les degrés alcooliques, ce qui montre que le degré de maturité du raisin doit, probablement, être un des facteurs déterminants affectant cette variable. On pourrait alors penser, parallèlement et en accord avec RIBÉREAU-GAYON (1976), que les maxima enregistrés dans le cas de la Basse Rioja pour les indices de Folin et les taux en matières réductrices s'expliquent par une plus grande durée de macération. L'extraction accrue des pentoses notamment, ainsi que celle des composés phénoliques en découle.

L'extrait sec d'un vin est en relation directe avec la teneur en sucres du moût original (AMERINE et OUGHT, 1980). C'est bien ce que l'on a constaté ici où les vins de la Basse Rioja ont les extraits secs les plus élevés. Les valeurs de ceux des vins de Haute Rioja sont à mettre en parallèle avec l'acidité totale élevée de ces mêmes vins (fortes teneurs

TABLEAU II

Comparaison des divers critères de sélection des variables.
Application aux trois groupes Haute Rioja, Basse Rioja, Rioja d'Alava

Critère de sélection	Ordre des variations choisies	Fiabilité du classement (%)
WILKS RAO	DEN, GRA, AV, AT, EXT, CEN, ACEN, MR, IF, ACT, ACL, K, GL, BUT, PR, N	90,37
	GRA, AT, EXT, ACEN, IF, ACL, K, BUT, PR, N	90,79
MINRESID	GRA, AT, EXT, ACEN, ACL, K, BUT, PR	93,42
	GRA, AT, EXT, K, BUT, PR	89,47
MAHAL	DEN, GRA, AV, AT, EXT, CEN, ACEN, MR, IF, ACT, ACL, K, GL, BUT, PR, N	90,37
	GRA, AT, EXT, ACEN, IF, ACL, K, BUT, PR, N	90,79
	DEN, GRA, AT, ACEN, K, BUT, PR	88,16
	GRA, AT, EXT, K, BUT, PR	89,47
MAXMINF	DEN, GRA, AV, AT, EXT, CEN, ACEN, MR, IF, ACT, ACL, K, GL, BUT, PR, N	90,37
	DEN, GRA, AT, ACEN, IF, ACL, K, BUT, PR, N	90,79
	DEN, GRA, AT, ACEN, K, BUT, PR	88,16
	GRA, AT, EXT, K, BUT, PR	89,47

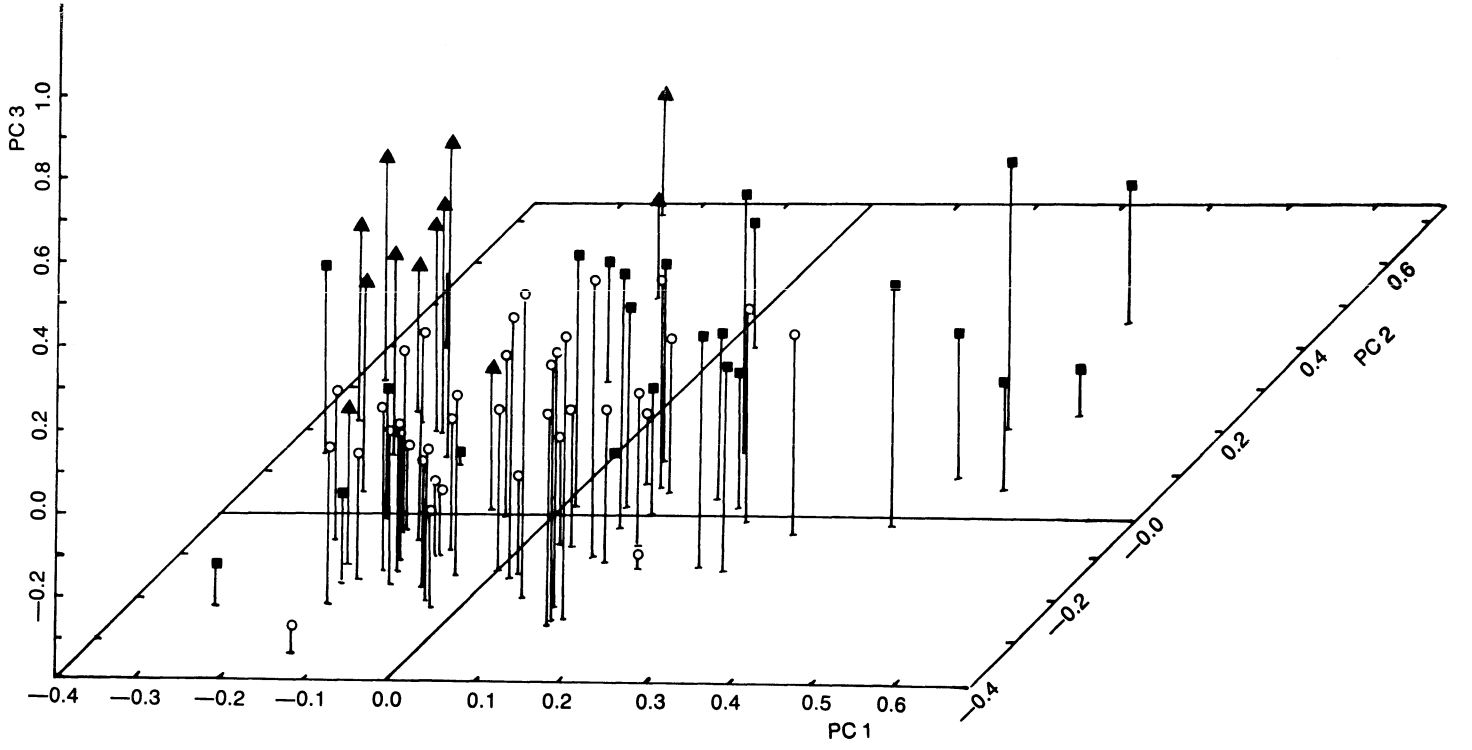


Fig. 1. — Représentation des vins considérés dans l'espace de trois premières composantes principales (■) Haute Rioja, (▲) Basse Rioja et (o) Rioja d'Alava.

relatives en acide tartrique). Il y a là accord parfait des données avec les facteurs climatiques et la variété des cépages, le Tempranillo donnant des moûts plus pauvres en sucres et plus acides que le Grenache.

L'analyse en composantes principales met en évidence les différences naturelles qui existent entre les vins des trois régions considérées. Ce genre d'analyse ne suppose a priori aucun classement de groupes, et permet d'observer le comportement général des données. La figure 1 montre la représentation des vins considérés dans l'espace des trois premières composantes principales. Quoique cette représentation contienne uniquement 51,4 p. 100 de la variance totale des données, on peut remarquer une nette tendance des vins à se grouper suivant la région d'origine.

Si l'on considère la matrice de corrélation obtenue (non représentée ici), on observe qu'il n'existe qu'une faible relation entre les paires de variables choisies : degré alcoolique et densité ($r = 0,62$), extrait sec et densité ($r = 0,59$), proline et degré alcoolique ($r = 0,68$), acide tartrique et acidité totale ($r = 0,65$), enfin potassium et cendres ($r = 0,66$). Le coœfficient de corrélation ne dépasse jamais 0,68. A ce niveau, aucune variable n'est éliminée afin d'éviter les erreurs au maximum.

Cinq critères différents (NIE et *al.*, 1975) ont été choisis pour sélectionner les variables (Tableau II). Un pourcentage maximal de bonne attribution est obtenu avec la valeur de 93,42 p. 100 quand huit variables (degré alcoolique, acidité totale, extrait sec, alcalinité des cendres, acide lactique, potassium, butanediol-2,3 et proline) sont utilisées pour l'analyse discriminante. On observe que le pourcentage des bonnes réponses diminue si l'on augmente ou si l'on diminue le nombre de variables sélectionnées. Ceci montre qu'un excès d'information peut nuire à la classification en groupes.

Dans le tableau III, on trouve les valeurs des cœfficients standards qui appartiennent à chaque fonction discriminante calculée. On remarquera que le potasium, l'extrait sec, l'acidité totale et le degré alcoolique sont des variables à forte influence (comme on s'y attendait d'ailleurs) sur la fonction discriminante principale. De même, ces variables ont un poids important dans les trois premières composantes principales de la figure 1.

Dans le même tableau III, on indique également les valeurs à partir desquelles on peut évaluer la « signification » des fonctions. La première fonction discriminante décrit 75 p. 100 de la variance totale qui existe entre les variables. La deuxième décrit les 25 p. 100 restants.

Un deuxième critère permet d'évaluer les fonctions discriminantes Il s'agit d'évaluer la signification statistique appelée « lambda de Wilks » (Λ), parmi toute l'information discriminante qui n'est pas intervenue dans le calcul des fonctions précédentes. Une valeur élevée de « lambda » indique

deuxième fonction pour ne pas perdre d'informations précieuses.

qu'il y a peu d'information perdue. Dans le cas étudié, le « lambda de Wilks » a une valeur de 0,484. Il est alors évident qu'il faut calculer la

TABLEAU III

Paramètres des fonctions discriminantes dérivées pour les groupes Haute Rioja, Basse Rioja et Rioja d'Alava

Fonction	Variables							
	GRA	AT	EXT	ACEN	ACL	K	BUT	PR
1	—0,492	—0,565	—0,600	0,031	—0,384	0,845	0,349	0,685
2	—1,072	0,510	0,096	0,450	0,214	—0,390	0,430	0,461

Critères d'évaluation des fonctions discriminantes

Fonction	Valeur propre	Variance (%)	Corrélateur canonique
1	3,21	75,05	0,873
2	1,07	24,95	0,718

Fonctions séparées	Λ	X^2	Degrés de liberté	Signification
0	0,115	148,1	16	0,000
1	0,484	49,7	7	0,000

L'importance de ces deux fonctions peut être mesurée visuellement en représentant les échantillons étudiés suivant un espace discriminant bidimensionnel. Comme le montre la figure 2 on obtient une bonne séparation des trois groupes anlysés.

L'application des fonctions de classification dérivées de la matrice de covariance totale ainsi que des centres de gravité des trois groupes permet l'évaluation d'un pourcentage correct d'attribution (Tableau IV). Cependant, la valeur élevée est au-delà des espérances de classification exacte de groupes d'objets. En effet, ce pourcentage est le résultat d'une classification des objets qui sont intervenus pour le calcul des fonctions discriminantes. A travers lui, on obtient une véritable estimation de l'erreur d'attribution lorsqu'on divise le nombre total d'objets en deux sous-groupes ; le groupe d'échantillonnage et le groupe testé. Sur le premier groupe on

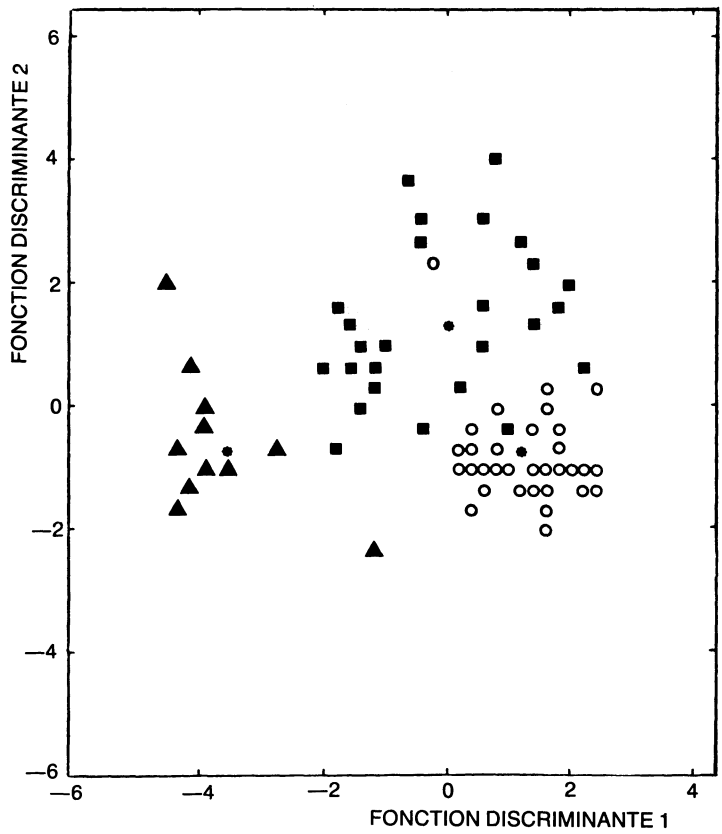


Fig. 2. — Représentation graphique des groupes Haute Rioja, Basse Rioja, Rioja d'Alava, en fonction de l'analyse discriminante.

(■) Haute Rioja
(▲) Basse Rioja
(o) Rioja d'Alava
(*) Centre de gravité du groupe étudié.

TABLEAU IV

Évaluation de l'erreur de classification. Groupes Haute Rioja, Basse Rioja et Rioja d'Alava.

Groupe	Nombre de vins	Groupe de classement obtenu par informatique		
		Haute Rioja	Basse Rioja	Rioja d'Alava
Haute Rioja	24	22	0	2
Basse Rioja	13	0	13	0
Rioja d'Alava	39	2	0	37

Nombre total de vins : 76
Pourcentage de vins correctement classés : 94,74 %
Erreur : 5,26 %

calcule les fonctions discriminantes et de classification. Ces fonctions sont évaluées grâce à la classification obtenue à partir du second groupe. Lorsqu'on recommence 26 fois le processus de fractionnement aléatoire des objets, en laissant environ 75 p. 100 de ces derniers dans le premier groupe, on obtient un pourcentage correct de prédiction voisin de 91,3 p. 100, avec un intervalle de confiance de 2,6 p. 100 au seuil de 0,5 p. 100.

TABLEAU V

Évaluation de l'aptitude à la prédiction des résultats dans le cadre de l'utilisation de la méthode de la machine d'apprentissage linéaire (LLM)

a) 8 variables choisies par le SLDA ont été utilisées pour le calcul

Haute Rioja - Basse Rioja	85,9 %	**Total : 87 %**
Haute Rioja - Rioja d'Alava	78,8 %	
Basse Rioja - Rioja d'Alava	96,3 %	

a : valeurs moyennes obtenues à partir d'un calcul sur 10 essais :
Groupe d'entraînement / Groupe d'essai.

Quand on applique la méthode de la machine d'apprentissage linéaire (LLM), basée sur une séparation des groupes à l'aide d'un hyperplan, on obtient les résultats du tableau V. Ces valeurs sont obtenues en effectuant 10 fois les calculs statistiques sur des groupes d'échantillonnage et de test, dont les populations respectives sont dans le rapport 4/1. Le pourcentage d'attribution correcte que l'on obtient (87 p. 100) est comparable à celui donné par la méthode d'analyse discriminante linéaire. Dans les deux cas, on observe une faible capacité prédictive lorsque les vins testés proviennent des régions les plus proches comme c'est le cas entre la Haute Rioja et la Rioja d'Alava.

Le tableau VI rassemble les résultats donnés par la méthode des K plus proches voisins (KNN). Après application de cette méthode sur les variables sélectionnées par l'analyse discriminante linéaire, on observe, qu'une augmentation du nombre de plus proches voisins entraîne une diminution du pourcentage de réussite. Cette tendance est évidente pour le groupe de la Haute Rioja, proche à bien des égards de celui de la Rioja d'Alava, au point que leurs « objets » sont confondus par l'analyse statistique. Ces erreurs de classement influent sur la valeur globale de la méthode. On obtient d'ailleurs des résultats analogues en traitant l'ensemble des variables avec cette méthode. Le pourcentage d'attribution correcte descend alors à 75,3 p. 100. D'un point de vue global, le résultat d'attribution correcte varie peu si au moyen des sous-programmes SELECT et

WEIGHT du logiciel ARTHUR on effectue une pondération de variance, dans le cadre d'une séparation en groupes des variables (82,3 p. 100 de réussite).

TABLEAU VI

Classsification des vins d'appellation Rioja à partir de la méthode des K plus proches voisins avec utilisation des variables sélectionées par l'analyse discriminante.

Groupe	Pourcentage correct de classification								
	K plus proche voisins								
	Objet	1-NN	3-NN	4-NN	5-NN	6-NN	7-NN	8-NN	9-NN	10-NN
Haute Rioja	24	66,7	54,2	50,0	45,8	54,2	45,8	45,8	45,8	29,2
Basse Rioja	13	92,3	84,6	84,6	92,3	92,3	92,3	92,3	92,3	92,3
Rioja d'Alava	39	92,3	97,4	92,3	97,4	97,4	100,0	100,0	100,0	100,0
		84,2	81,6	77,6	80,3	82,9	81,6	81,6	81,6	76,3

Pourcentage total de réussite* : 80,9

* Valeur moyenne calculée à partir du rapport du nombre total d'objets bien attribués au nombre total d'objets.

CONCLUSIONS

Au cours de ce travail, les techniques d'analyse des données ont été utilisées pour rechercher, de façon objective, l'origine de vins provenant d'appellations de caractéristiques voisines, en utilisant des paramètres physico-chimiques facilement accessibles. Bien que ceux-ci puissent être individuellement, entachés d'une certaine probalité d'erreurs, l'analyse multivariante minimise ces erreurs en faisant intervenir l'ensemble des mesures.

Dans le cadre de cette étude, l'utilisation d'une méthode de classification basée sur l'attribution à un groupe à l'aide d'une mesure de distance entre objets (méthode KNN) conduit à des pourcentages de bonne attribution inférieurs à ceux qu'on obtient par analyse statistique discriminante ou par régression linéaire. Les résultats obtenus ici sont toutefois moins nets que ceux de TAPIAS (1985), qui portaient sur des vins issus de régions très différentes les unes des autres. De plus, l'échantillonnage choisi dans chaque cas n'a pas la même importance. Il est probable que la prise en compte d'autres variables, en plus de celles étudiées (comme des données spectroscopiques (densités optiques) ou chromatographiques) pourrait affiner les résultats (MORET et *al.*, 1984 ; FRANK et KOWALSKI, 1984).

Remerciements.

Les auteurs remercient bien sincèrement les docteurs ès Sciences F. ILLAS et J. RUBIO pour leur collaboration efficace dans l'utilisation de l'analyse des données.

Ce travail a été subventionné par la CAICYT, organisme du Ministère de l'Education et de la Science (projet n° 2838/83), à qui nous exprimons ici toute notre gratitude.

Manuscrit reçu le 6 juin 1986 ; accepté pour publication le 12 novembre 1986.

RÉSUMÉ

L'application de l'analyse multidimensionnelle des données à la reconnaissance des vins de trois appellations de la Rioja, permet de choisir huit variables physico-chimiques, facilement accessibles, comme étant hautement significatives. En plus de ces paramètres analytiques, le cépage et les données climatiques ont un rôle important. Les meilleurs résultats sont obtenus au moyen de la méthode d'analyse discriminante linéaire des données avec laquelle le pourcentage d'attribution correcte des vins atteint 91,3 p. 100.

SUMMARY

Application of multidimensional data analysis to the recognition of three Rioja appellation wines led to selecting 8 easly-obtained physicochemical variables as being highly significant. In addition to these analytical parameters, variety and climatic conditions play an important role. The best results were obtained using linear discriminant analysis of the data, which gave 91,3 p. 100 correct recognition of the wines.

ZUSAMMENFASSUNG

Die Anwendung vielschichtiger Analysen bei Gegebenheiten von Weinen drei verschiedener Qualitäts und Herkunftsbezeichnungen der Rioja. erlaubt es, acht physich-chemische Veränderlichleiten zu wälhen, die leicht Zugänglich und doch hoch bezeichnend sind Zusätzlich zu diesen Unbeständigktein Spielt die Analyse der Rebsorte und der Klimatischen Vorraussetzungen eine wichtige Rolle; die (Ergebnisse die am besten zufriedenstellem) werden mit einer geradlinigen Analysemethode erziett, bei hierbei erreicht die korrekte Zuteilung der Weine 91,3 p. 100.

RESUMEN

La aplicación de métodos de análisis multivariante en el reconocimiento de vinos pertenecientes a las tres zonas de Rioja, permite elegir ocho variables físico-químicas, fácilmente obtenibles, como las más importantes para la clasificación correcta de las muestras. Factores que influyen decisivamente en la diferenciación de los vinos producidos son el distinto tipo de variedad y el clima. El porcentaje máximo de predicción correcta obtenido alcanza el valor de 91,3 p. 100 si la clasificación se lleva a cabo aplicando el método de análisis discriminante estadístico lineal.

RIASSUNTO

L'applicazione dell'analisi multidimensionale dei dati al riconoscimento dei vini di tre demominazoni controllate della Rioja permette di scegliere otto variabili fisico-chimichi, facilmente accessibili come altamente significativi. In più di questi parametri analitici, il vitigno e il clima hanno un ruolo importante. Il migliori risultati sono ottenuti col metodo di analisi discriminante lineare dei dati con la quale il percentuale d'attribuzione correetta dei vini raggiunge 91,3 p. 100.

REFERENCES BIBLIOGRAPHIQUES

AMERINE M.-A. et OUGH C.-S., 1980. Methods for analysis of must and wines. John Wiley & Sons. New York.

BERMEJO F. et BALUJA SANTOS C., 1974. Determinacion de litio en los ácidos aminopolicarboxílicos como reguladores espectrofotométri-acidos aminopolicarboxílicos como reguladores espectrofotométricos. *Acta Científica Compostelana, XI,* (1), 3-12.

BERTUCCIOLI M., CLEMENTI S. et GIOLIETTI G., 1984. Relazione fra dati sensoriali ed analitici dei vini. *Vini d'Italia,* **3**, 27-36.

FRANK I.-E. et KOWALSKI B.-R., 1984. Prediction of wine quality and geographic origin from chemical measurements by partial least-squares regression modeling. *Anal. Chim. Acta,* **162**, 241-251.

COOMANS D., MASSART D.-L. et KAUFMAN L., 1979. Optimization by statistical linear discriminant analysis in analytical chemistry. *Anal. Chim. Acta.* **112**, 97-122.

HARPER A.-M., DUEWER D.-L., KOWALSKI B.-R. et FASCHING J.-L., 1977. Arthur and experimental data analysis: the heuristic use of polyalgorithm. *In:* Chemometrics Theory and Applications., B.-R. Kowalski (ed.)., pp 14-52, American Chemical Society, Washington, D.C. Programa de Calculo Arthur disponible de Informetrix, Inc, P.O. Box 25808, Seattle, WA 98125, USA.

INIGUEZ M., 1980. Estudio de las relacions enológicas sobre vinos de Rioja. Instituto de Estudios Riojanos. Logrono.

MASSART D.-L. et KAUFMAN L., 983. The interpretation of analytical chemical data by the use of cluster analysis. John Wiley & Sons. New York.

MEDINA B. et VAN ZELLER A.-M., 1984. Différenciation des vins de trois régions de France. *Connaissance Vigne Vin,* **18**, N° 4, 225-235.

MORET I., SCARPONI G. et CESCON P., 1984. Aroma components as discriminating parameters in the chemometric classification of Venetian white wines. *J. Sci. Food Agric.,* **35**, 1004-1111.

NIE N.-H., HULL C.-H., JENKINS J.-G., STEINBRENNER K. et BENT D.H., 1975. Statistical package for the social sciences pp 434-467. 2nd ed. McGraw-Hill. New York.

NOBLE A.-C., WILLIAMS A.-A. et LANGRON S.P., 1984. Descriptive analysis and quality ratings of 1976 wines from four Bordeaux communes. *J. Sci. Food Agric.*, **35**, 88-89.

POLO M.-C., MARTIN CORDERO P. et CABEZUDO M.-D., 11984. Influence des caractéristiques variétales de moûts de cépages différents sur la fermentation alcoolique par une seule souche de levure sélectionnée, *Bull. O.I.V.*, **638**, 312-321.

POLO M.-C., HERRAIZ M. et CABEZUDO M.-D., 198. A study of nitrogen fertilization and fruit maturity as an approach for obtaining the analytical profiles of wines and wines groupes. Instrumental analysis of foods and beverages: recent progress by G. Charalambous and G. Inglett. Academic Press. **2**, 357-374.

RAPP A. et GUNTERT M., 1985. Contribution to wine characterization of the grape variety White Riesling. 2. Investigation on the volatiles of German white wines of the varieties White Riesling, Muller-Thurgan and Sylvaner. *Vitis*, **24**, 139.

RIBÉREAU-GAYON J., PEYNAUD P. et RIBÉREAU-GAYON P., 1976. Sciences et techniques du vin, Tome 1, pp 290. Dunod, Paris.

SHINOHARA T., 1984. Importance of volatile substances of wine. Formation and effects on quality. *Bull. O.I.V.*, **57**, (641-642), 606-618.

TAPIAS R.-M., LARRECHI M.-S., GUASCH J., RUBIO J. et RIUS F.-X., 1986. Enological parameters and pattern recognition methods in the geographic differentiation of Spanish red wines. *Am. J. Enol. Vitic.*, **37**, n° 3, 195-201.

VAN DER VOET H., DOORNBOS D.-A., MEEMS M. et HAAR G. VAN DE., 1984. The use of pattern recognition techniques in chemical differentiation between Bordeaux and Bourgogne wines. *Anal. Chim. Acta*, **159**, 159-171.

VARMUZA K., 1980. Pattern Recognition in Chemistry. Springer-Verlag. New York.