

Differentiation between Wines Originating from Different Red Wine Cultivars and Wine Regions by the Application of Stepwise Discriminant Analysis to Gas Chromatographic Data

J. MARAIS, P. C. VAN ROOYEN and C. S. DU PLESSIS;

Oenological and Viticultural Research Institute, Private Bag X5026, Stellenbosch, 7600.

Stepwise discriminant analysis (SDA) was applied to gas chromatographic data of some volatile compounds extracted with freon from red wines. By means of this statistical method, Pinotage and Cabernet Sauvignon wines (1978 vintage) could be classified into separate groups, the most effective variables being hexanol and i-amyl acetate. Pinotage wines from three, and Cabernet Sauvignon wines from two localities (1979 vintage) were subjected to classification according to origin. Successful classifications were obtained using the variables ethyl acetate, ethyl caprate, i-amyl acetate as well as a compound tentatively identified as i-valeric acid.

With the increasing sophistication and application of analytical techniques such as gas chromatography, analytical information tends to proliferate to such an extent that interpretation by visual examination of the data becomes extremely difficult and in many cases, almost impossible. Several statistical methods have been used in recent years in an attempt to determine which components affect the quality of various products, such as foodstuffs, alcoholic and non-alcoholic beverages.

A number of research workers reported on the evaluation of quantitative analytical data by stepwise discriminant analysis (SDA), the aim being to differentiate between specified groups or individual samples. Powers & Keith (1968) applied SDA to classify potato chips and coffee samples. Discriminant techniques have also been used successfully on blends of Pepsi-Cola and Coca-Cola (Young, Bargmann & Powers, 1970). In the brewing industry, computer identification and classification of different types of beers, as well as beers from different geographical regions, were demonstrated with the aid of SDA (Brown, Clapperton & Dalgliesh, 1974; Helbert & Hoff, 1974; Reiner & Piendl, 1974; Hoff, Helbert & Chicoye, 1975; Brown & Clapperton, 1978).

In wine research SDA has successfully been applied by Schreier *et al.* (1976) who differentiated between various muscat cultivars using the quantitative distribution of terpene compounds. Similarly Rapp *et al.* (1978) differentiated between Riesling and Morio Muscat, as well as between Riesling, Sylvaner and B 6-18 (B) using SDA on 27 aroma compounds. Wines and brandies were also classified successfully according to vintage and regions by Cordner *et al.* (1978), Kwan & Kowalski (1978) and Schreier & Reiner (1979).

The aim of this study was to identify factors responsible for differences between wines from different cultivars and from different regions, as well as to guide further investigations to the selection and identification of wine components, which can be used as quality predicting parameters.

MATERIALS AND METHODS

Material: Forty-six dry red wines made from the Pinotage cultivar (Pinot noir × Cinsaut) from the Stellenbosch,

Durbanville and Worcester areas, as well as 32 dry red wines made from the Cabernet Sauvignon cultivar from the Stellenbosch and McGregor areas, were used in this study. Wines from both the 1978 as well as the 1979 vintage were used.

Grapes were harvested periodically over approximately six weeks in various stages of ripeness from the same vineyards to induce variations in quality. The wines were made according to standard procedures used at the experimental cellar of the institute.

Gas chromatography: The wines were extracted with freon and analysed quantitatively for esters and higher alcohols by means of gas chromatography (Marais & Houtman, 1979). In order to limit the possible effects of time on wine composition, all the samples were extracted and analysed within one month. A Hewlett Packard 5840A gas chromatograph fitted with an automatic peak integrator was used.

SDA: The BMD07M computer program (Dixon, 1968) was utilized, performing a multiple discriminant analysis in a stepwise manner. By entering one variable at each step the data are searched to select the variable of greatest discriminatory value subsequent to which a classification is made. This is repeated until all the variables have been tested. Variables can subsequently be ranked according to their efficiency as discriminants. The program also calculates canonical variables which, after plotting, yields a two-dimensional representation of the dispersion of the samples around their respective average response values.

RESULTS AND DISCUSSION

A typical Pinotage wine chromatogram is illustrated in Figure 1.

The components determined, as well as their average concentrations for the Pinotage and Cabernet Sauvignon wines respectively, are given in Table 1.

In the Pinotage wines, ethyl caprate (peak 32, Fig. 1) was inseparable from another component (peak 33, Fig. 1), which, by spiking the solutions with the respective compounds, has tentatively been identified as i-valeric acid. Therefore, peaks 32 and 33 were measured together

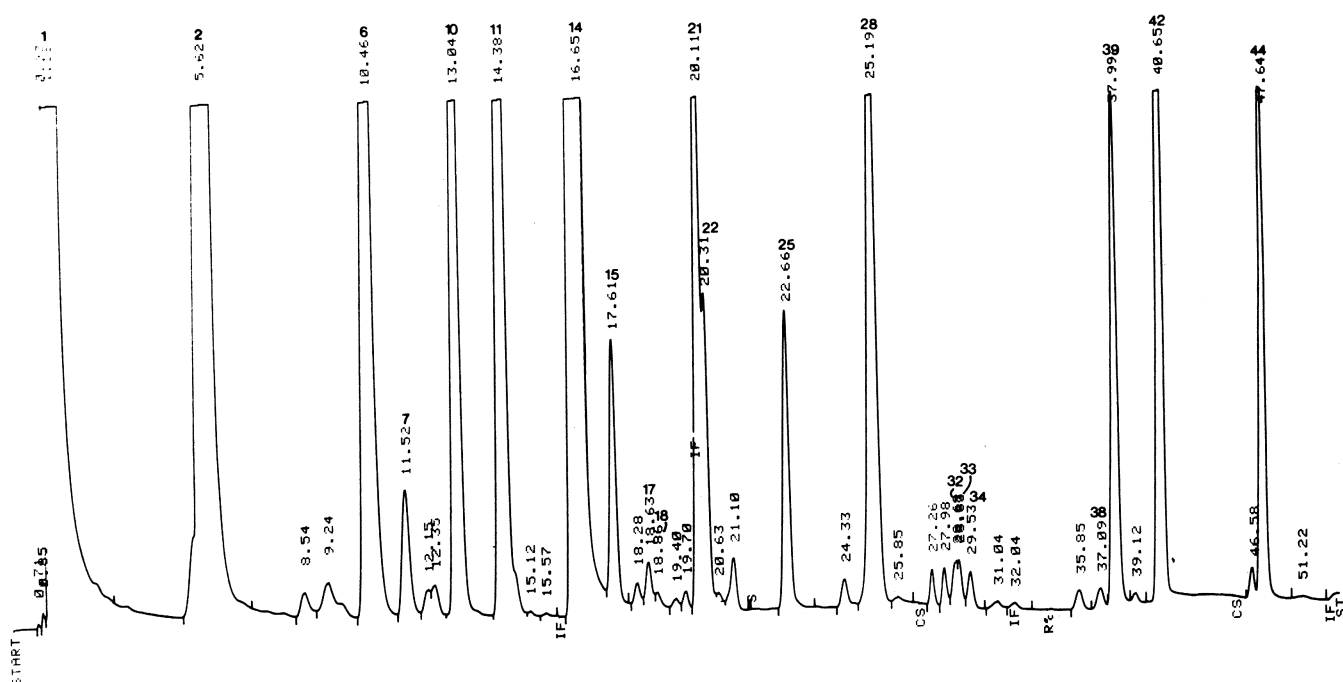


FIG. 1

Chromatogram of Freon 11 extract of a Pinotage wine (attenuation 2^{10}). 1 = Freon 11; 2 = Ethyl acetate; 6 = tert.—Amyl alcohol (internal standard); 7 = Ethyl butyrate; 10 = i-Butanol; 11 = i-Amyl acetate; 14 = Amyl alcohols; 15 = Ethyl caproate; 17 = Hexyl acetate; 21 = Ethyl lactate; 22 = Hexanol; 25 = Ethyl caprylate; 28 = Ethyl nonanoate (internal standard); 32 = Ethyl caprate; 33 = i-Valeric acid; 34 = Diethyl succinate; 38 = 2-Phenethyl acetate; 39 = Hexanoic acid; 42 = 2-Phenyl ethanol; 44 = Octanoic acid.

TABLE 1
Average concentrations of esters, higher alcohols and acids of Pinotage and Cabernet Sauvignon wines (1978 and 1979 vintages).

Vintage	Pinotage		Cabernet Sauvignon	
	1978	1979	1978	1979
<i>Esters (mg/l)</i>				
Ethyl acetate (2) ^a	64,99	70,66	32,74	42,00
Ethyl butyrate (7)	0,75	0,88	0,31	0,19
i-Amyl acetate (11)	6,32	4,85	1,99	0,75
Ethyl caproate (15)	0,59	0,63	0,34	0,26
Hexyl acetate (17)	0,08	0,05	—	—
Ethyl lactate (21)	2,38	5,88	7,88	10,87
Ethyl caprylate (25)	0,76	0,94	0,53	0,53
See footnotes	0,44 ^b	0,33 ^b	0,65 ^c	0,47 ^c
Diethyl succinate (34)	0,28	0,80	1,34	3,28
2-Phenethyl acetate (38)	0,22	0,13	0,14	0,05
<i>Higher alcohols (mg/l)</i>				
i-Butanol (10)	46,86	41,22	55,45	46,43
Amyl alcohols (14)	197,32	200,57	328,90	327,73
Hexanol (22)	0,66	0,78	1,39	1,70
2-Phenyl ethanol (42)	13,48	16,69	47,11	70,90
<i>Acids (mg/l)</i>				
Hexanoic acid (39)	3,48	4,48	2,33	1,93
Octanoic acid (44)	3,79	3,79	2,46	1,24

a = Peak No., Fig. 1

b = Peak 32 + 33, noted as ethyl caprate

c = Peak 33, i-valeric acid

and the values are identified as ethyl caprate. In the Cabernet Sauvignon wines the concentration of ethyl caprate was insignificant, whereas the concentration of i-valeric acid was so much higher than in Pinotage wines, that it can probably be assumed that mainly i-valeric acid was measured.

Three sets of data were used in all the analyses viz. (1) four higher alcohols (peaks 10, 14, 22 and 42); (2) ten esters (peaks 2, 7, 11, 15, 17, 21, 25, 32, 34 and 38); and (3) the total of the first two sets plus two additional acids (peaks 39 and 44). Data processing was executed on a Burroughs 7700 computer.

Classification according to cultivar: SDA was applied to the data from 10 Pinotage and 10 Cabernet Sauvignon wines from the Stellenbosch area (1978 vintage). Successful separations were achieved by using any of the three sets of variables. The two most important variables obtained from the analysis of each set are given in Table 2.

TABLE 2

Two most important discriminant variables for the classification of Pinotage and Cabernet Sauvignon wines using set 1 (higher alcohols), set 2 (esters) and set 3 (higher alcohols + esters + acids).

	Variables	F-value to enter
Set 1	Hexanol	125,45*
	Amyl alcohols	18,18*
Set 2	i-Amyl acetate	207,80*
	Ethyl butyrate	13,80*
Set 3	i-Amyl acetate	207,80*
	Octanoic acid	10,36**

* = Highly significant at $P \leq 0,01$

** = Significant at $P \leq 0,05$

The results obtained for the second and third sets must be treated with caution, because the ratios of number of wines to the number of variables are not sufficiently high. A canonical plot of the two variables of the first set is given in Figure 2, illustrating that of the four higher alcohols examined, (i-butanol; hexanol; 2-phenyl ethanol; amyl alcohols) hexanol and amyl alcohols are sufficient for a complete separation of the two cultivars.

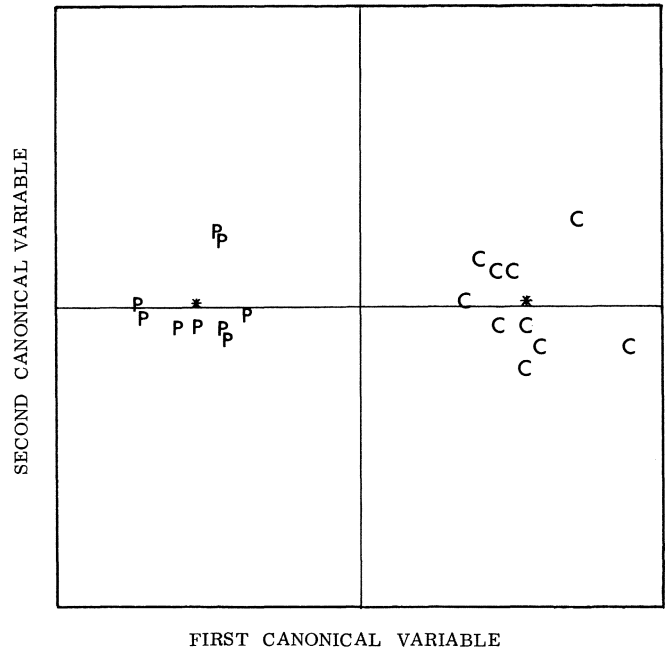


FIG. 2

A canonical plot showing separation of Pinotage and Cabernet Sauvignon wines from the Stellenbosch area on the basis of four higher alcohols (1978 vintage).

P = Pinotage
 C = Cabernet Sauvignon
 * = Mathematical centre of each group.

For sets 2 and 3, bearing in mind the number of wines to variables, it was found that i-amyl acetate offered the best discriminatory power as a single variable, indicating a relative constant difference between the two cultivars for all the wines used in the analysis.

Classification according to region: Pinotage: In this investigation 36 Pinotage wines, originating in equal numbers from three regions, viz. Stellenbosch, Durbanville and Worcester (1979 vintage), were subjected to SDA. Employing the same three sets of variables described previously, the most successful separation was obtained with set 3, which contained all the variables. Table 3 contains a summary of the classification results.

TABLE 3

Classification of 36 Pinotage wines into three regions employing three sets of variables.

Region	Set 1 (Higher alcohols)			Set 2 (Esters)			Set 3 (All the variables)		
	S	D	W	S	D	W	S	D	W
S	9	1	2	8	2	2	11	1	0
D	1	9	2	2	10	0	0	12	0
W	3	2	7	1	0	11	0	0	12

S = Stellenbosch
 D = Durbanville
 W = Worcester

Verwag men die dat Eth. caprylate in beter skeiding sal veroorsaak nie

From the above results it can be deduced that a virtually complete separation could be obtained by using all the variables. Higher alcohols alone were not very successful as discriminants, especially for the Worcester Pinotage wines, where 40% of the wines were classified into the other two regions. When all the variables are used (set 3), the order of importance of these discriminants confirm the above conclusion (Table 4). In this analysis i-butanol emerges as the third most important discriminant, however, the other higher alcohols play a much lesser role. The canonical plot employing the two most important variables of set 3 is given in Figure 3.

TABLE 4

Discriminant variables of set 3 in decreasing order of importance for Pinotage wines from three regions

Variable entered	F-value to enter
Ethyl caprate	8,19*
Ethyl acetate	6,15*
i-Butanol	5,96*
i-Amyl acetate	5,17*
Diethyl succinate	4,42*
Hexyl acetate	3,92*
Octanoic acid	3,77*
Hexanol	3,41*
Ethyl caprylate	2,57**
2-Phenyl ethanol	1,98
2-Phenethyl acetate	0,94
Ethyl caproate	0,84
Ethyl lactate	0,55
Ethyl butyrate	0,53
Hexanoic acid	0,31
Amyl alcohols	0,21

* = Highly significant at P ≤ 0,01
 ** = Significant at P ≤ 0,05

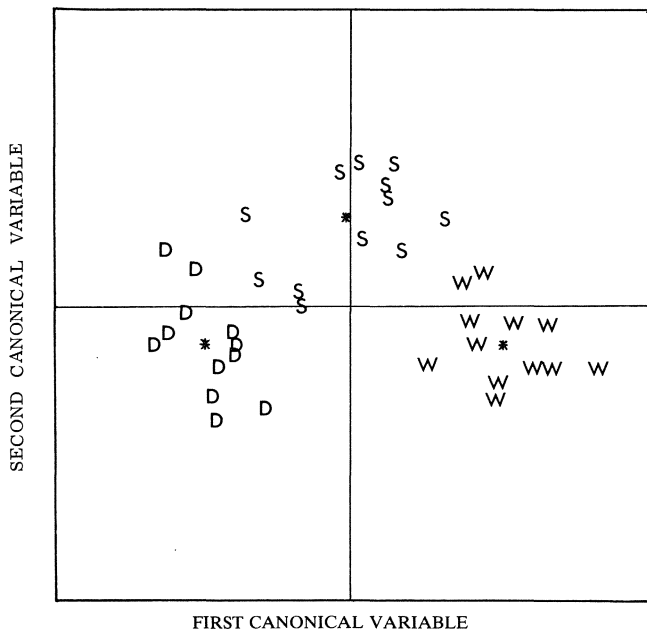


FIG. 3

A canonical plot showing separation of Pinotage wines from Stellenbosch, Durbanville and Worcester areas on the basis of four higher alcohols, ten esters and two acids (1979 vintage).

S = Stellenbosch
 D = Durbanville
 W = Worcester
 * = Mathematical centre of each group.

It is clear from Figure 3 that the wines from the three regions could be separated quite successfully by using mainly ethyl caprate and ethyl acetate. It follows, therefore, that these two variables, as well as the other highly significant discriminants, are affected by pedo-climatic factors. Similar results were obtained in a separate SDA analysis in which only the eight highly significant discriminants were used (in order to have fewer variables in the equations).

Cabernet Sauvignon: In the case of Cabernet Sauvignon, SDA was applied to data from wines from two regions for the 1979 vintage, viz. 10 and 12 wines from the Stellenbosch and McGregor areas, respectively. The most successful separations were obtained employing variables in sets 2 and 3. The resulting classifications are given in Table 5.

TABLE 5

Classification of 22 Cabernet Sauvignon wines into two regions employing three sets of variables.

Region	Set 1 (Higher alcohols)		Set 2 (Esters)		Set 3 (All the variables)	
	S	M	S	M	S	M
S	6	4	10	0	10	0
M	5	7	0	12	0	12

S = Stellenbosch
 M = McGregor

The discriminant components listed in Table 6, in order of importance for sets 2 and 3, reveal that i-valeric acid is, in both cases, the most effective discriminant for Cabernet Sauvignon wines from the two regions. In the case of set 2 (esters), i-amyl acetate and ethyl butyrate are also highly significant discriminants. When set 3 is used, i-butanol takes second place, and i-amyl acetate becomes insignificant. In a separate SDA analysis using only i-valeric acid, i-butanol, i-amyl acetate and ethyl butyrate as discriminant variables, an excellent separation according to region was obtained, with F-values of 26,47**; 9,12**; 1,20 n.s. and 0,31 n.s., respectively. A canonical plot of the separation of the wines from the two regions employing all the variables given in Table 6, is depicted in Figure 4.

CONCLUSIONS

The general applicability of SDA for the investigation of pedo-climatic effects on wine composition is demonstrated. Regional differentiation of Cabernet Sauvignon wines could be obtained by using relatively few fermentation-produced flavour compounds. Classification of Pinotage wines according to region proved to be more difficult, necessitating the utilisation of more variables.

The usefulness of SDA in this respect does not merely lie in the identification of wines from different regions, or in the identification of cultivars from gas chromatographic analyses, but rather in its value as an aid to the identification of quality and character parameters which are affected by factors such as pedo-climatic, viticultural or cultivar differences. This can lead to a better understanding of ways to positively influence wine quality and yield technical evidence for the delimitation of "wines of origin" regions. Other pattern recognition methods, such as those used by Kwan, Kowalski & Skogerboe (1979) should also be employed in further studies.

TABLE 6

Discriminant variables of sets 2 and 3 in decreasing order of importance for Cabernet Sauvignon wines from two regions.

Set 2		Set 3	
Variable entered	F-value to enter	Variable entered	F-value to enter
i-Valeric acid	26,47*	i-Valeric acid	26,47*
i-Amyl acetate	6,35*	i-Butanol	9,12**
Ethyl butyrate	4,64*	Ethyl acetate	6,31
Ethyl caproate	3,88**	Ethyl butyrate	5,56
Unknown (Peak 18)	3,39**	Ethyl caprylate	3,69
2-Phenethyl acetate	2,94**	Hexanol	3,47
Ethyl acetate	2,87**	Amyl alcohols	3,32
Ethyl lactate	0,50	Diethyl succinate	3,26
Diethyl succinate	0,42	2-Phenyl ethanol	1,71
		Ethyl caproate	1,68
		Octanoic acid	1,44
		Ethyl lactate	0,40
		Hexanoic acid	0,30
		i-Amyl acetate	0,16
		Unknown (Peak 18)	0,07
		2-Phenethyl acetate	0,04

* = Highly significant at $P \leq 0,01$
 ** = Significant at $P \leq 0,05$

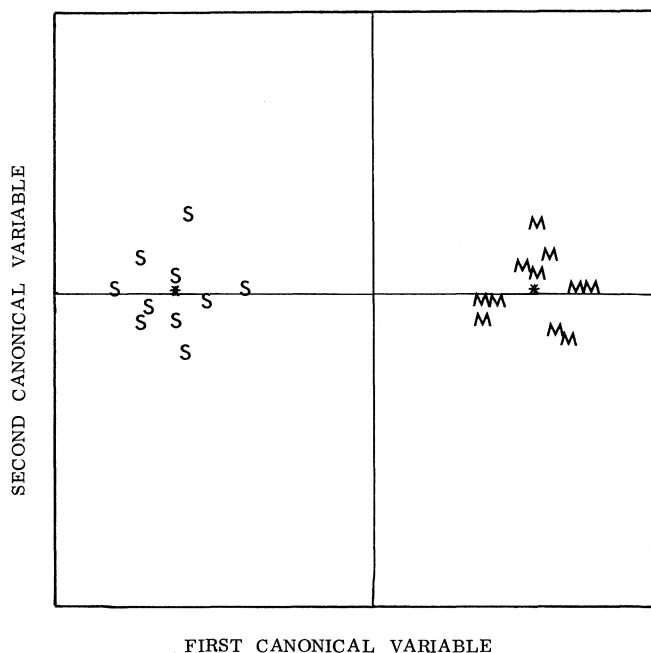


FIG. 4

A canonical plot showing separation of Cabernet Sauvignon wines from Stellenbosch and McGregor areas on the basis of four higher alcohols, nine esters and three acids (1979 vintage).

S = Stellenbosch
 M = McGregor
 * = Mathematical centre of each group.

LITERATURE CITED

BROWN, D. G. W., CLAPPERTON, J. F. & DALGLIESH, C. E., 1974. The language of flavor and its use in product specification. *Am. Soc. Brew. Chem. Proc.* **32**, 1-4.
 BROWN, D. G. W. & CLAPPERTON, J. F., 1978. Discriminant analysis of sensory and instrumental data on beer. *J. Inst. Brew.* **84**, 318-323.

CORDNER, C. W., OUGH, C. S., KASIMATIS, A. N. & KISSLER, J. J., 1978. Effects of crop level on chemical composition and headspace volatiles of Lodi Zinfandel grapes and wines. *Am. J. Enol. Vitic.* **29**, 247-253
 DIXON, W. J. (ed.), 1968. BMD Biomedical Computer Programs, University of California Press, Berkeley.
 HELBERT, J. R. & HOFF, J. T., 1974. Discriminant analysis—A powerful tool for brewery research. *Am. Soc. Brew. Chem. Proc.* **32**, 43-49.
 HOFF, J. T., HELBERT, J. R. & CHICOYE, E., 1975. Classification of lager beers by computer analysis of volatile profiles. *MBAA Tech. Quart.* **12**, 209-213.
 KWAN, W. O. & KOWALSKI, B. R., 1978. Classification of wines by applying pattern recognition to chemical composition data. *J. Food Sci.* **43**, 1320-1323.
 KWAN, W. O., KOWALSKI, B. R. & SKOGERBOE, R. K., 1979. Pattern recognition analysis of elemental data. Wines of *Vitis vinifera* cv. Pinot noir from France and the United States. *J. Agric. Food Chem.* **27**, 1321-1326.
 MARAIS, J. & HOUTMAN, A. C., 1979. Quantitative gas chromatographic determination of specific esters and higher alcohols in wine using freon extraction. *Am. J. Enol. Vitic.* **30**, 250-252.
 POWERS, J. J. & KEITH, ELIZABETH S., 1968. Stepwise discriminant analysis of gas chromatographic data as an aid in classifying the flavor quality of foods. *J. Food Sci.* **33**, 207-213.
 RAPP, A., HASTRICH, H., ENGEL, L. & KNIPSER, W., 1978. In "Flavor of Foods and Beverages". Academic Press N. Y., S. F., Lond.
 REINER, L. & PIENDL, A., 1974. Zur Unterscheidung von Biersorten mittels der Diskriminanzanalyse. *Brauwissenschaft* **27**, 33-39.
 SCHREIER, P., DRAWERT, F., JUNKER, A. & REINER, L., 1976. Anwendung der multiplen Diskriminanzanalyse zur Differenzierung von Rebsorten an Hand der quantitativen Verteilung flüchtiger Weinhaltstoffe. *Mitt. Klosterneuburg* **26**, 225-234.
 SCHREIER, P. & REINER, L., 1979. Characterisation and differentiation of grape brandies by multiple discriminant analysis. *J. Sci. Food Agric.* **30**, 319-327.
 YOUNG, L. L., BARGMANN, R. E. & POWERS, J. J., 1970. Correlation between gas chromatographic patterns and flavor evaluation of chemical mixtures and of cola beverages. *J. Food Sci.* **35**, 219-223.