Wine and quality. A cluster analysis to assess certain representative aspects relating to the quality

Stefania Chironi[1], *Marzia Ingrassia.*
University of Palermo.

Abstract

This research, in which the technique of Cluster Analysis is used, will provide information on how wine producers can evaluate certain characteristics (variables) of their product more widely associated as indicative of quality. The analysis identified homogeneous groups containing within them other, related variables, which can be considered similar. Those attributes which are by themselves most representative of the product's quality are also highlighted. Finally, the cluster analysis made it possible to comment on the behaviour of producers, and looking at how they seek to meet market expectations based on their knowledge of customer's habits in the buying and consuming of wine.

JEL classification: C39, M31, Q13.
Keywords: Wine, quality indicators, Cluster analysis.

[1]Corresponding author: chirstef@unipa.it, Dep. E.S.A.F. (Economia dei Sistemi Agro Forestali), Faculty of Agricolture, University of Palermo, Viale delle Scienze, Building n. 13, Palermo, Italy. This paper is a result of the Research Programme "(ex. 60%)" "Awareness of connections between agricultural products and territories", financed by the University of Palermo, directed by Prof. Stefania Chironi. It was realized by Prof. S. Chironi (Chapter 3 and Conclusions) and Dr. M. Ingrassia (Introduction and Chapter 2).

1 Introduction

The issue of quality is now one of the main talking points for all who work in the agricultural chain of production or in the agri-food industry. Quality is now an essential attribute for any agro-food product and the EU stressed its importance in the Common Agricultural Policy (CAP). Indeed, according to regulations; Reg. EEC 2092/914[2], Reg. EEC 2081/92[3] and 2082/92[4] Reg EEC - which are considered the CAP's three 'pillars' of quality - it was to be identified with a product's traceability and geographical origin. This was to strengthen ties with the area of production, promoting the use of traditional working methods and environmentally-friendly practices.

From the consumer's perspective, however, the perception of quality is extremely subjective and, it is identified, in food industry products particularly, with attributes often linked to a product's inherent properties, its taste, freshness or genuineness. But for the producer the concept of quality is most often associated with the strictness of production, aspects of the production process, food safety, health and environmental protection, accurate information for consumers, higher prices and more profit.

With regards to wine in particular, the concept of quality assumes more complex levels of meaning, combining the organoleptic characteristics of quality, with those linked to the territory of origin, and those related to external aspects such as the label (with other information contained in it) or the name of the producer- all elements that may influence the consumer at the time of purchase by creating 'guarantees' of quality.

This work, adding to a previous paper[5], aims to identify, through the use of the technique of cluster analysis, which, according to wine producers, are the homogeneous groupings of product characteristics (variables) that can be considered similar and which are interrelated. The technique has allowed us to highlight those attributes which, by themselves, are most representative of the product quality.

Twenty one elements (variables) were identified as indicative of the wine's quality, and a survey of the 21 selected variables was carried out on a panel of wine producers. Thanks to the clustering technique it was possible to get some ideas about what, according to producers, are the qualitative factors considered most adapted to the customer's expectations (Customer satisfaction) and to the

[2]Reg. EEC 2092/91 regarding biological methods of production of agricultural products.

[3]Reg. EEC 2081/92 regarding protection of geographic indication and denomination of origin of agricultural and agro-food products.

[4]Reg. EEC 2082/92 regarding certification of agricultural and agro-food products.

[5]Chironi, (2006). Perception and transfer of information on wine quality. The opinion of producer and consumer.

market.

The results of the clustering are critical to identify groups of variables whose degree of similarity can help in distinguishing the elements that connote quality wine.

The implementation of the cluster analysis to the results - obtained through the administration of a similar questionnaire to a sample of consumers - was moved to another location.

2 Method

2.1 The sample scheme

The work, as already mentioned in the introduction, is an extension of an earlier investigation. For our analysis it was desirable to start with the Palermo and Trapani wine producers, who, having participated in Vinitaly in 2006, would have an interest in the creation of quality wines and the expansion of their market share.

The "population" of participating producers in Vinitaly 2006 (over the area Palermo and Trapani) was, from official sources, equivalent to N = 107[6].

Since the population from which the sample was to be extracted was already stratified as follows: N1 = 74 for Trapani and N2=33 for Palermo (N = 107), it was possible to apply 'stratified proportional' sampling, allowing us to increase the accuracy without increasing the size of the sample. They were therefore considered as two layers of the population from which the desired samples were extracted, in proportion to the size of the layer,

Since the sample size was n = 62, a sample size of n1 = 43 were randomly extracted in proportion from the 1st, N1 = 74 layer, and a sample size of n2 = 19 from the second, N2 = 33 layer, obtaining a total sample size, n1 + n2 = n = 62.

2.2 The questionnaire

For the interviews was a questionnaire carried out containing a list of 21 quality variables, held by us, as mentioned earlier, as characterizing the evaluation of the 'wine product', to which the interviewee was to give a score (1 to 4) according to its importance for them. The selection of variables was carried out in order to apply a hierarchical Cluster Analysis on the producers' answers, and then analyse these variables into groups and between groups.

The 21 variables listed on the survey given to producers are:

[6]Catalogue of the Regional Institute of Vine and Wine for the Vinitaly 2006.

1. the shape of the bottle, 2. attractive label, 3. label presence, 4. wine colour, 5. alcohol, 6. scent, 7. vintage, 8. organoleptic characteristics, 9. brand, 10. organic grapes, 11. bottlers mark, 12. cork bottle-stop, 13. silicone cork, 14. designated origin, 15. awards received, 16. varying price ranges, 17. grape origin, 18. wine store sale, 19. large scale retail sales, 20. production diversification, 21. limited bottle supply.

The variables relate to both 'external' aspects which attract the consumer visually or be 'elements of distinction' for the producer - such as the bottle shape, label, graphics and information provided on the back - and the wine's 'intrinsic' qualities, like the colour, the grape source, alcohol levels, the smell and the year of harvest (vintage). Specific aspects of the market are also highlighted such as the availability of the product at wine retailers or large-scale distributors, the name of the bottler and the company brand.

2.3 The method of analysis

Cluster analyses, as is known, are used when faced with the problem of creating relatively homogeneous groups in a set of variables. It is a multivariate analysis technique through which it is possible to combine statistical data, so as to minimize the 'logical separation' between statistical data inside each group and to thus maximize it between groups. The 'logical distance' is quantified by measures of defined similarity/dissimilarity between the data. So cluster analysis puts the elements of a certain set together into non-predefined groups (clusters), so that those belonging to the same group are as homogeneous as possible, relative to all the characteristics of each element, and elements belonging to different groups are heterogeneous.

Clustering, unlike other analysis techniques, can throw light on a set of homogeneous groups in *the absence of any knowledge on the shape and the number of groups to be obtained.*

The source data is more important in this type of analysis than in any other, as are the notions of *distance* and *similarity* (similar, but opposite concepts): distance corresponds to a similarity. Several methods are used to calculate the distance between two or more variables.

The 'Euclidean' distance is often calculated, despite having the disadvantage of depending very strongly on the units of measure and this creates a difference in distance if the values are equivalent but measured differently. This problem is resolved by a standardizing of the variables, and thus obtains comparable variables and distances. In our case the variables have not been standardized because they used a single scale of measurement.

From the various existing methods, we considered it appropriate to apply the *technique of hierarchical grouping* using the method of *complete linkage* which lends itself to the formation of clusters where the elements are *naturally* forming, taking into account their unique characteristics and different *'blocks'*. According to this method, the distances between clusters are determined on the basis of the greatest distance between two elements belonging to different clusters.

The cases and clusters unite on the basis of criteria to be adopted at every step of the aggregation. These criteria are based on the *matrix of distances*, in particular the maximum distance shows the minimum similarity. The method of complete linkage defines distance between two clusters between their further cases.

In our case, the matrix of distances produced by the twenty one 'quality variables' was determined on the basis of 'Euclidean distance'.

The graph, called 'dendrogram' (Fig.1) is the main way of representing the output from hierarchical cluster analysis. It makes it possible to view the stages of aggregation effectively, showing the level of cluster aggregation according to increasing order.

Vertical lines in the dendrogram indicate the union of two clusters, while the positions of those lines in the scale of reference indicate the distance at which these clusters aggregate[7].

For a more accurate assessment of the process of aggregation, in addition to the dendrogram the table contains an '*agglomeration program*' (Table 1) which may be studied. Or the complete and detailed sequence of the aggregation process, with the codes for the cases of step by step aggregation of clusters (fused clusters), the distance in aggregation 'coefficients', corresponding to the maximum distance between cases which can form a cluster, the '*stage*' in which the listed case has already been involved in an aggregation (the last stage shows the highest degree of aggregation) and the 'successive stages' in which the formed cluster will again be subject to aggregation.

2.4 Results

The results obtained are interesting in that variables with different characteristics and meanings have been grouped together, starting from groups with higher degree of homogeneity and descending to the lower degrees until we reach two major clusters (1 -10 and 1 -6), of all variables considered. Producers gave a score to each of the variables in their answers to the survey and the collected data was then placed in tables and prepared as input for the factorial method.

[7]Distances, for representation scope, are measured assuming (SPSS) a value between 1 (min distance) and 25 (max distance).

The results of the analysis are shown in Table 1, which shows the 'agglomeration program' and the 'dendrogram' is also very useful because it shows, quite clearly, the hierarchy of gradually forming groups and the number of clusters to be considered, and it shows which elements belong to each cluster.

In particular, from the dendrogram analysis and Table 1, five major groups can be identified, which decrease during the last stage (stage 20), to the largest cluster (1 and 6). This final cluster includes within it all other cluster generated using the method and the distance (coefficient of cluster 1-6) between the two variables is therefore increased. The hierarchy of the five major clusters was spread thus: Cluster 1 -6, Cluster 1 -10, Cluster 1 -2, Cluster 1-13, Cluster 1-3.

An analysis of the dendrogram identifies at an early stage those clusters consisting of characteristics that are most alike, which show a great homogeneity in the allocation of judgement from the wine producers, continuing on to more advanced stages we gradually come to ever larger clusters containing an ever increasing number of features and characterized by higher 'coefficients' indicating a lesser degree of similarity between characteristics (Table 1).

3 Results discussion

In light of results obtained through the use of the tool it becomes possible to make some observations on what wine producers understand as quality in wine. They put themselves with the buyer of the product in some cases, the consumer - whilst also trying to respect, as much as possible, the decisions of the business, based in all cases on maximizing profit.

At stage 1 there is the first cluster (16 -18), identifiable as 'different price ranges' and 'specialist wine retailer sales.' They are by far the most similar variables in the sense that the sample winemakers' judgements in terms of contribution to quality were distributed with the same percentages, we can therefore determine that the producers hold these two aspects as equally representative of quality. Following the hierarchical pattern (table 1), at stage 2 we find cluster (6 -8) containing the variables 'scent' and 'organoleptic characteristics', where again we find that producers have given similar weightings to these two factors. Indeed it can be seen that producers have given the highest rating to both variables, as they are distributed with the same percentage, among the highest possible.

The producer considers these as the most defining in the expression of quality - seeing things from the perspective of consumers who, through their sense of smell can classify the product: smelling a wine's fragrance is the first thing anyone does before taking a drink!

The scent and organoleptic characteristics are well-paired, defining as a whole

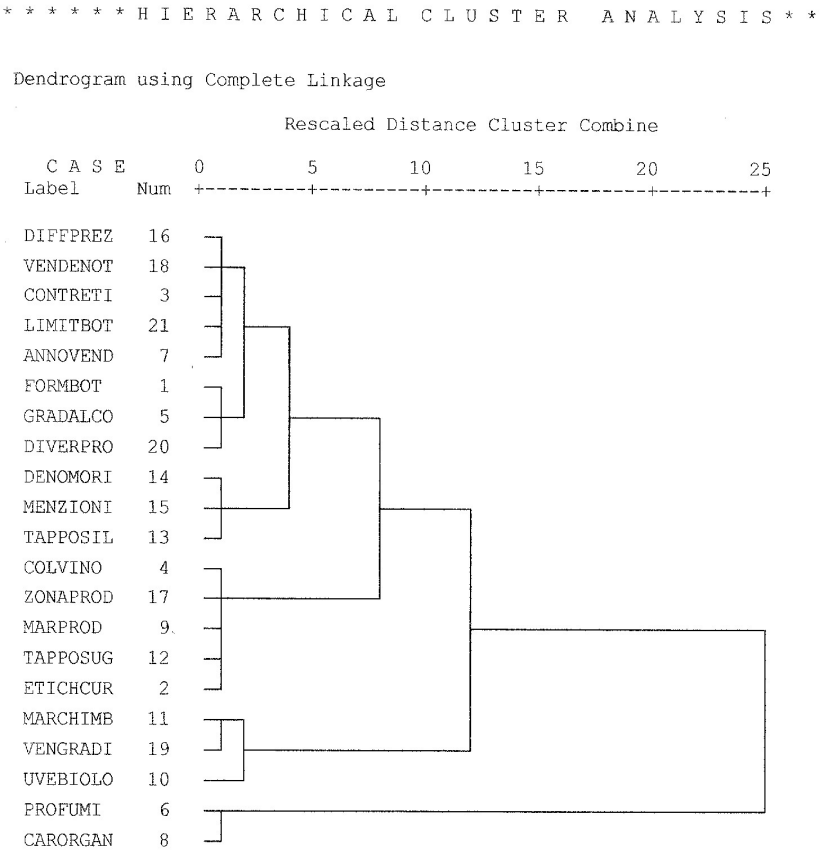
Table 1: Agglomeration Program

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	16	18	.010	0	0	12
2	6	8	.010	0	0	20
3	1	5	5,120	0	0	7
4	3	21	15.360	0	0	8
5	4	17	20.480	0	0	13
6	9	12	35.840	0	0	9
7	1	20	36.810	3	0	15
8	3	7	46.730	4	0	12
9	2	9	66.560	0	0	13
10	14	15	67.860	0	0	14
11	11	19	83.210	0	0	16
12	3	16	192.660	8	1	15
13	2	4	255.060	9	5	18
14	13	14	300.180	0	10	17
15	1	3	390.430	7	12	17
16	10	11	419.860	0	11	19
17	1	13	978.600	15	14	18
18	1	2	2352.760	17	13	19
19	1	10	3456.840	18	16	20
20	1	6	7592.190	19	2	0

everything that serves to express the original bond with the site of production, which strongly characterizes quality today. Continuing on to an analysis of the next stage, we have cluster (1-5) - 'shape of the bottle' and 'alcohol'. We are still at a level where the producer continues to see a strong similarity between two variables.

At the next cluster (3-21) - 'presence of labels and 'limited number of bottles' the producers are beginning to express slight differences in their judgements which creates a greater dissimilarity between characteristics within the cluster. At the following stage cluster (4 -17) is detected - 'colour of the wine' and 'production area of the grapes'. The wine producer clearly perceives a strong correlation between these two features since they indicate and distinguish the quality of one wine from another and as such are considered two very important factors by the producers.

Figure 1: Dendrogram



Clusters (9-12) - 'brand producers' and 'cork bottle-stop', (1-20) 'bottle shape' and 'production diversification', (3-7) 'presence of labels' and 'vintage' produced a similar coefficient level, indicating less homogeneous groupings than in previous clusters. They were nonetheless found to be important variables contributing to the quality of a wine.

It is also noted that cluster (1-20) is generated by adding variable 20 - 'production diversification' - to cluster (1-5) and similarly cluster (3 -7) is generated by adding variable 7 - 'vintage' - to cluster (3-21). Continuing with the hierarchical process, we see three clusters bearing similar coefficients and thus containing analogous variables, they are; cluster (2-9) 'attractive label' and 'brand', cluster (14 -15) - 'designated origin' and 'awards received' and cluster (11-19) - 'bottlers

mark' and 'large scale retail sales'. Cluster (2-9) is obtained by adding variable 2 - 'attractive label' to cluster (9-12). This means that producers found variable 2 very similar to those making up cluster (9-12). Therefore 'attractive label', 'producers brand' and 'cork bottle-stop' are all still very similar variables and remain very important factors for producers (who gave these variables a very high rating in the questionnaire, with a similar percentage) along with 'wine colour' and 'grape origin' that make up cluster (2-4).

Hierarchically below come clusters (14-15) 'designated origin' and 'awards received', (11-19) 'bottlers mark' and 'large scale retail sales' and (3-16) 'presence of label' and 'varying price ranges'. This last cluster is at a much higher hierarchical level than the first cluster (16-18), in fact it is generated by adding variable 3 - 'label presence '.

In the next stage come clusters (13 -14) 'silicone cork' and 'designation of origin' and the first large cluster, cluster (1-3) 'bottle shape' and 'label presence'. These clusters are formed with variables that are less similar from the perspective of the producers compared with clusters identified in the earlier stages. In fact, (Table 1) shows that the coefficients attributed to these variables are higher (greater distance/less similarity). This is an indication that in attributing a judgement on these variables, although the producers have recognised how each is representative of wine quality, they did not express a consensus, the score being distributed on all levels. Obviously, these are still all fairly important factors for producers. At the next stage is cluster (10-11) 'organic grapes' and 'bottlers' mark'. At this level there are ever more marked distances between variables in the cluster, compared to the earlier stages, and as we head towards clusters with larger coefficients, the producer perceives increasing differences between the inherent variables of the group, not considering them indicative of the same level of quality.

It is interesting to note how distances begin to increase at stage 17, because it is more difficult for the producer to distinguish the similarities between the variables.

Cluster (1-13) contains the 'bottle shape' and 'cork bottle-stop', features that are not considered similar by producers in their ratings of quality, but rather indicative of different qualitative levels.

We can furthermore note that producers today are particularly careful in finding bottles that have the most pleasing shape to consumers (it has been attributed of high to average importance), while not identifying the presence of a real cork bottle-stop as indicative of quality (defined as being of unimportant to average importance). This is due to the fact that often the producer can hardly separate the quality from the cost of production and so considers unimportant the presence of a cork stopper, rather than a silicone one.

It can also be seen that cluster (13-14) - 'cork bottle-stop' and 'designated origin', is connected to cluster (1-13) 'bottle shape' and 'cork bottle-stop', thereby coming together to form the second largest single group of homogeneous variables, that includes 'bottle shape', 'alcohol', 'production diversification', 'designated origin', 'awards received' and 'silicone cork'. Indeed these characteristics (although we must remember that we are already at the stage where only small similarities exist), are grouped together in logical terms, because if a producer gives special attention to a product in terms of all these elements of quality then they will even more gratified if it wins awards.

At the next stage is the third major group, identified by cluster (1-2) 'bottle shape', and 'attractive label', this cluster is included within cluster (1-20) bottle shape', 'alcohol' and 'production diversification', cluster (13 -14) - 'silicone cork', 'awards received' and 'designated origin', and finally cluster (2-4), including 'attractive label', 'cork bottle-stop', 'brand', 'grape origin' and 'wine colour'.It is also noted how, as clusters are get larger, they include a growing number of variables which, due to their number, cannot have the same levels of similarity to those grouped in the early stages, but which continue to be more 'logically' similar to each other, compared to the those not included in the grouping. The fourth major group is represented by cluster (1 -10) 'bottle shape' and 'organic grapes', this cluster includes the cluster (1-13) 'bottle shape', 'alcohol', 'production diversification', 'designated origin', 'awards received' and 'silicone cork', (4-2)[8] 'wine colour', 'grape origin', 'brand', 'cork bottle-stop' and 'attractive label', and the cluster (10-11)[9] 'organic grapes', 'large scale retail sales' and 'bottlers' mark'.

The last large group finally, that encompasses all the precedents, is the cluster (1-6) 'bottle shape' and 'scent. As can be seen clearly from the dendrogram this cluster encompasses the entire hierarchy generated by the application of the method that identifies the various 'bunches' of clusters linked with each other 'logically'.

Finally, it is interesting to note that the cluster (6-8) 'scent' and 'organoleptic characteristics', remains separated from the others and forms a separate group (as can be seen clearly on the dendrogram).

4 Conclusions

The analysis of variables obtained through the application of the cluster method, has enabled us to comment on wine producers' behaviour with regards to their perception of quality.

[8]Inside cluster (1 - 2).
[9]Inside cluster (1 – 10).

The producer as already pointed out, will, in certain respects, have a different concept of quality,to the consumer, because quality, for the producer also means reaching a good level of production from a business that will enable it to achieve adequate income and this is not always in line with what the consumers' expectations might be.

The application of the cluster method has created 5 large, homogenous groups that each encompass different, but interrelated variables and considered similar for certain characteristics. The cluster of 'scent' and 'organoleptic characteristics' deserves special attention, they are the attributes that received the highest rating - showing how for producers they are the only characteristics that are the most representative of the quality of a wine.

Characteristics, therefore, very similar to each other but 'unlike' all the others, exactly because they are so clearly indicative of the quality of this type of product. As seen, these characteristics in fact, remain independent of all other factors as the producers are aware that the scent of a wine and its characteristics are derived from elements that underlie the quality of a wine (land, type of grape, etc.). Therefore these are 'intrinsic' characteristics for producers, key elements in recognizing and distinguishing a quality wine.

Within this first major order of clustering is included a number of variables that include, on the other hand, a number of 'extrinsic' characteristics of the product, but which are nevertheless relevant mainly because they help the brand identification of a specific company that provokes familiarity for the consumer.

Another cluster binds a whole series of features on which there is significant agreement between producers, such as production diversification, the linking to the designated origin, grape origin which becomes highly significant if one considers the close link with the territory and a differentiation of quality desired by legislation. This is extended with limited edition bottles and awards that a product that has certainly received attention in its production process wins; and thanks also to the awards, may sway the consumer in their final choice.

The producer considers them as aspects with which to express quality, seeing it through the eyes of consumers who, through many of these elements can recognise the quality of the product.

Anticipating even for a comparison, some aspect of the foregoing analysis for the consumer, for example, we note that the consumer considers the 'cork bottle-stop' most qualifying, unlike the producer who, as we have seen, makes no particular distinction with silicone, perhaps unable to separate the influence it has on the cost of production.

Moreover, 'organic grapes', 'designated origin' and 'wine retailer sales' are strongly qualifying variables for the consumer while they are not identified as

similar in expressing the quality of wine for the producers.

Finally, it is also interesting to note that although variables 'varying price ranges' and 'wine retailer sales' are considered by producers as equally representative of the wine quality - they are aspects about which the consumer is very much interested.

The Cluster Analysis has put together a number of issues relating to quality and made it possible to examine how the differences between producers and consumers are in some cases significant and, therefore, pose a good sign for the producer to understand that orienting their production decisions to the market doesn't always mean going along with the company's goals of maximizing profits.

References

Cicchitelli, G., Herzel, A. and Montanari, G. (1992). Il campionamento statistico. Bologna: Il Mulino.

Chironi, S. (2006). La percezione e la trasmissione delle informazioni sulla qualità di un vino. Il punto di vista del consumatore e quello del produttore.

Corbetta, P.(1992). Metodi di analisi multivariata per le scienze sociali. Bologna: Il Mulino.

Corbetta, P. (1999). Metodologie e tecniche della ricerca sociale. Bologna: Il Mulino.

Fabbris, L. (1997). Statistica Multivariata. Milano: Mc. Graw Hill.

Vianelli, S. and Ingrassia G. (2000). Istituzioni di Metodologia Statistica. Palermo: Palumbo.