APPLICATION OF DISCRIMINANT ANALYSIS IN THE CHARACTERIZATION OF PHYSICOCHEMICAL DATA OF BUFFALO MILK

Asif Mahmood, Sumaira Naeem
University of Gujrat, Department of Chemistry, Gujrat, Pakistan
E-mail: asifmahmood023@gmail.com

Abstract
Milk is a very important food. Physical and chemical quality of milk is affected by addition of water and ice. Milk quality monitoring program that provides a representative and reliable estimation of the quality of milk has become an important necessity. A comprehensive monitoring program provides large meaningless data. In this research work discriminant analysis (DA) was used to analyze the physicochemical data of milk of buffalo's dairy farms of Gujranwala, Pakistan. Parameters were divided into three groups, physical (pH, Electrical conductivity, specific gravity, titratable acidity, total solid), chemical (Ash, Fat, Protein, Lactose) and minerals (Na, Ca, Mg, K, Fe). Discriminant analysis indicates that only 57.1% of parameters were correctly classified. This is due the overlapping nature of parameters. Discriminant analysis (DA) identified ten significant parameters (Ca, Na, Mg, Fe, titratable acidity (TTA), Fat, specific gravity (SG), Ash and K which discriminate (distinguish) the milk quality of seven dairy farms. Discriminant analysis classified seven dairy farms 100.0% correctly. The result of classification shows that there are significant differences between the quality of milk of seven dairy farms, which are expressed by six discriminate functions. This study illustrates the benefit of discriminant analysis for interpreting complex data sets in the analysis of spatial variations in milk quality, and to plan for future studies.

Keywords: discriminant analysis, milk, physicochemical analysis, buffalo, Gujranwala

Submitted: 16.05.2011 Reviewed: 15.06.2011 Accepted: 25.06.2011

1. INTRODUCTION

Milk is secreted by all species of mammals to supply nutrition and immunological protection to the young. It performs these functions with a large array of distinctive compounds (Jenness et al., 1999). It contains a wide range of dietary components of vital importance like water, proteins, lactose, minerals and vitamins. These components are necessary for normal growth and performing different functions for the body systems. Milk is valuable diet for the expectant mothers and growing children. Milk has been recognized as nature's single most complete food (Shah and Khan, 1982). Unfortunately, due to unorganized and non-regulated marketing system, the quality of milk is hardly maintained at consumer level. Physical and chemical quality of milk is affected by addition of water and ice, which is a common practice in Pakistan. Addition of water and ice changes ratio of different constituents. Raw milk supply to Gujranwala takes place from dairy farms in the surrounding areas, and is based on unorganized selling system. Hence, milk for consumption is hardly assumed to be of high quality.

Milk quality monitoring is very important. A monitoring program that provides a representative and reliable estimation of the quality of milk has become an important necessity. Consequently, comprehensive monitoring programs that include frequent milk sampling at numerous sites and that consists a full analysis of a large number of physicochemical parameters are to be designed for proper management of milk quality. Real physicochemical data are mostly noisy, it means that they are not normally distributed, often co-linear or autocorrelated, containing outliers or errors etc. In of order to avoid this problem discriminant analysis was used.

Discriminant analysis has been used to distinguish the milk and cheese of various species (Fresno et al., 1995; Herrero-Martinez et al., 2000; Martin-Hernandez et al., 1992;

Rodriguez et al., 1999). It has been used to distinguish between two different diets (Favretto, Vojnovic and Campisi, 1994), to distinguish the different physiological conditions of the animals and the different season based on the metabolic profile (Biagi et al., 1991). It has also been used to identify preventively those subjects which were about to give birth, according to milk composition (Harwood et al., 1991).

The objective of this research work was to:

1. Classified the parameters
2. Determinate which variables are the most useful for discriminating between dairy farms
3. Classified the dairy farms

Brief review of discriminant Analysis used in this study

Discriminant analysis is a technique for classifying a set of observations into predefined classes. It operates on raw data and the technique constructs a discriminant function for each group (Lattin, Carroll and Green, 2003; Wunderlin, 2001). A simple linear discriminant function transforms an original set of measurements on a sample into a single discriminant score (Sanchez Lopez et al., 2004). DA involves the determination of a linear equation that will predict which group the case belongs to. The form of the equation or function is

$$D = v_1 X_1 + v_2 X_2 + v_3 X_3 ....v_i X_i + a$$

Where D = discriminate function

v = the discriminant coefficient or weight for that variable

X = respondent's score for that variable

a = constant

i = the number of predictor variables

The statistical software packages SPSS 16 and Statgraphic were used for the discrimnant analysis.

2. MATERIALS AND METHODS

Collection of samples

Forty nine samples of raw buffalo milk were collected from seven dairy farms (seven from each). All the samples were brought to Laboratory of Chemistry Department, University of Gujrat, Gujrat, Pakistan. Physicochemical analysis of milk samples was carried out in laboratory.

Physical analysis

The pH was measured using a digital pH-meter ((InolabWTW Series 720) calibrated with pH 4 and 7 buffers. Electrical conductivity (EC) was measured using conductivity meter (SP 536). pH and electrical conductivity were determined at the time of sample collection. Specific gravity (SG) was determined using pycnometer as described by AOAC (AOAC, 2000). Total solids (TS) content was determined by gravimetric method as described by AOAC (AOAC, 2000). Titratable acidity (TTA) was determined by titrimetric method as described by AOAC (AOAC, 2000).

Chemical analysis

Ash content was determined by gravimetric method using a muffle furnace at 550 ^{0}C as described by AOAC (AOAC, 2000). Fat content was determined by Rose-Gottlieb method as described by AOAC (AOAC, 2000). Protein content was estimated by formal titration method (Davide, 1977). Lactose content was determined by using Fehling's solution method (Triebold, 2000).

Minerals analysis

For minerals analysis, 50 ml of each milk sample was taken in beaker. Concentrated HNO_3 and $HClO_4$ were added in 1: 1 ratio. Mixture was slowly digested on hot plate almost to dryness. Beaker was removed from hot plate and H_2O_2 was added and again heated on hot plate until solution became colourless. Digested sample was cooled, filtered and quantitatively transferred to a flask. Solution was diluted to 100 ml with deionized double distilled water (Perveen, Mumtaz and Usmani, 2005). Concentration of Na and K was determined using flame photometer (Cole-Parmer Flame Photometer EW-83055-05). Concentration of Ca, Mg and Fe was determined using Atomic absorption spectrophotometer (Perkin Elmer 2380).

3. RESULTS AND DISCUSSION

Fourteen physicochemical parameters were determined during this study. Descriptive statistics of all the parameters is given in Table 1.

Classification of parameters

Parameters were divided into three groups, physical **(**pH, EC, SG, TTA, TS), chemical (Ash, Fat, Protein, Lactose) and minerals (Na, Ca, Mg, K, Fe). Standard mode discriminant analysis was used to check the correctness of classification.

Table 1. Descriptive Statistics of milk quality of seven dairy farms of Gujranwala

Parameter	Unit	Minimum	Maximum	Mean	Std.Deviation
pH		5.49	7.00	6.47	0.35
EC	uS/cm	5.35	8.01	6.64	0.68
SG	%	1.024	1.035	1.028	0.002
TTA	%	0.11	0.29	0.19	0.04
TS	%	13.21	20.18	17.58	1.41
Ash	%	0.37	0.98	0.58	0.15
Fat	%	6.15	9.21	7.62	0.83
Protein	%	2.95	4.97	4.02	0.50
Lactose	%	3.40	7.11	5.17	0.94
Na	mg/l	12.96	36.76	23.27	6.92
Ca	mg/l	45.06	80.95	62.00	10.52
Mg	mg/l	9.86	18.16	13.61	2.27
K	mg/l	1.23	4.53	2.41	0.80
Fe	mg/l	0.49	3.31	1.76	0.86

Table 2. Classification of parameters

Parameter	Actual Group	Predicted Group	P value	Squared Mahalanobis Distance to Centroid
pH	Physical	Physical	0.949	0.104
EC	Physical	Chemical**	0.581	1.087
SG	Physical	Physical	0.940	0.124
TTA	Physical	Physical	0.910	0.189
TS	Physical	Physical	0.065	5.470
Ash	Chemical	Physical**	0.969	0.062
Fat	Chemical	Physical**	0.825	0.385
Protein	Chemical	Minerals**	0.697	0.722
Lactose	Chemical	Chemical	0.631	0.920
Na	Minerals	Minerals	0.385	1.907
Ca	Minerals	Minerals	0.651	0.857
Mg	Minerals	Minerals	0.552	1.190
K	Minerals	Chemical**	0.585	1.071
Fe	Minerals	Chemical**	0.901	0.209

**. Misclassified case   a. 57.1% of original grouped cases correctly classified.

Only 57.1% of original grouped cases correctly classified as shown in Table 2. In first group electrical conductivity (EC) was classified incorrectly into chemical group. EC of milk is due metals in ash. Ash and fat of second group were classified incorrectly into first group. Ash of milk is mainly due to total solid (TS). Due to this reason ash falls in physical group. Specific gravity (SG) is an index of fat. Due to this reason fat falls in physical group. Protein was classified incorrectly into third group. In milk approximately 67% of the calcium, 35% of the magnesium bound within the casein micelle and the remainder are soluble in the serum phase. In third group K and Fe were classified incorrectly into second group. In figure 1 plot of scores of two functions shows that three groups of parameters are overlapping to each other. This indicate that there no distinguishable boundary between three groups. Parameters of different groups depend on each other.

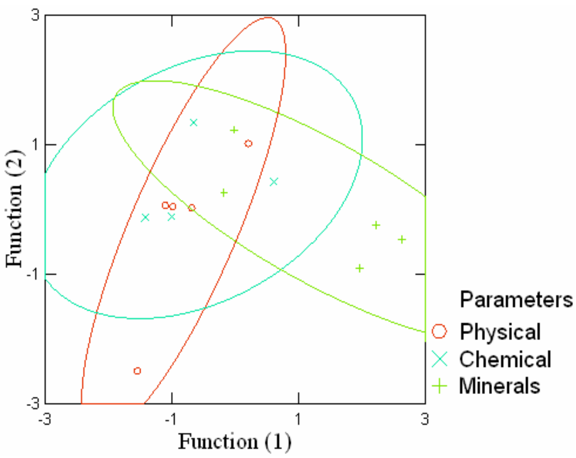

Fig.1. Plot of scores of two functions

Classification of dairy farms

Discriminant analysis was used to find functions (linear combinations) of the observed data (called discriminant functions) that best separate the milk quality of seven farms. Standard mode discriminant analysis was applied in present study. DA was applied on raw data. Seven discriminate functions (DFs) were found to discriminate the seven dairy farms as shown in Table 3.

Table 3. Eigen-values of discriminant function for milk quality of seven dairy farms

Function	Eigen value	% of Variance	Cumulative %
1	64.285	57.4	57.4
2	25.718	23.0	80.3
3	14.938	13.3	93.7
4	3.887	3.5	97.1
5	1.985	1.8	98.9
6	1.231	1.1	100.0

Wilk's Lambda test showed that all the functions are statistically significant (Table 4). Six DFs explained 57.4%, 23.0%, 13.3%, 3.5%, 1.8% and 1.1% of the total spatial variance respectively. The relative contribution of each parameter to six functions is given in Table 5.

Table 4. Wilks' Lambda test of DFs for spatial variation of milk quality of seven dairy farms

Test of Function(s)	Wilks' Lambda	Chi-square	Sig.
1 through 6	0.000	514.332	0.000
2 through 6	0.000	357.628	0.000
3 through 6	0.002	234.428	0.000
4 through 6	0.031	130.601	0.000
5 through 6	0.150	71.104	0.000
6	0.448	30.092	0.000

In first function Ca and Na exhibited strong contribution in discriminating the seven dairy forms and account for most of the expected spatial variations in the quality of milk of seven dairy forms, while less contribution exhibited from other parameters. In second function, contribution of Mg in explaining the spatial variations is major as shown in Table 5.
In third function Fe is most discriminating parameter. TTA is most discriminating parameter in fourth function. Fat and SG are major discriminating parameters in fifth function. In sixth function Ash and K are most discriminating parameters. Lactose is less discriminating parameter as indicated by lower score in all functions. Lactose plays a major role in milk synthesis.

Table 5. Discriminant function coefficients of spatial variation of milk quality of seven dairy farms

	Function					
	1	2	3	4	5	6
pH	0.056	0.312	0.344	-0.137	0.260	-0.212
EC	0.115	0.012	0.102	0.085	0.304	-0.034
SG	0.381	-0.232	0.458	-0.452	**0.624**	-0.204
TTA	-0.151	-0.519	-0.064	**0.535**	-1.239	0.028
TS	-0.151	0.113	0.268	0.184	0.291	-0.097
Ash	-0.140	0.450	-0.381	0.381	0.413	**0.910**
Fat	0.296	0.112	0.404	-0.075	**0.629**	-0.174
Protein	0.369	0.000	0.210	-0.230	0.383	-0.541
Lactose	0.313	0.038	-0.375	0.290	-0.127	0.038
Na	**0.644**	-0.759	0.293	0.367	0.249	-0.177
Ca	**0.933**	0.430	-0.401	0.155	-0.126	0.000
Mg	-0.027	**0.502**	-0.094	-0.421	-0.015	0.143
K	0.357	0.003	0.461	-0.774	-0.209	**0.506**
Fe	0.296	0.389	**1.161**	0.103	0.110	-0.066

Parameters were grouped based on function coefficients and following functions are indicated:
Function 1: Ca, Na
Function 2: Mg
Function 3: Fe
Function 4: TTA
Function 5: Fat, SG
Function 6: Ash, K

Table 6. Classification results for discriminant analysis of seven dairy farms

Dairy form	% correct a	Predicted Group Membership						
		1	2	3	4	5	6	7
1	100.0	7	0	0	0	0	0	0
2	100.0	0	7	0	0	0	0	0
3	100.0	0	0	7	0	0	0	0
4	100.0	0	0	0	7	0	0	0
5	100.0	0	0	0	0	7	0	0
6	100.0	0	0	0	0	0	7	0
7	100.0	0	0	0	0	0	0	7

It is the major osmole in milk and the process of synthesis of lactose is responsible for drawing water into the milk as it is being formed in the mammary epithelial cells. Because of the close relationship between lactose synthesis and the amount of water drawn into milk, lactose content is the least variable component of milk.

The classification matrix showed that 100.0 % of the cases are correctly classified to their respective groups, as shown in Table 6.

The result of classification shows that there are significant differences between milk quality of seven dairy farms, which are expressed by six discriminate functions.

This study provides important information about milk quality of buffalo's dairy farms, which can be used to plan future studies with minimum parameters, samples and with lower cost.

4. CONCLUSION

DA indicates that predefined groups of parameters are classified 57.1% correctly. This is due to overlapping nature of parameters. DA allowed a reduction in the dimensionality of the large data set and indicated a few significant parameters responsible for large variations in milk quality that could reduce the number of sampling parameters. Hence, this study illustrates that discriminant analysis is an excellent exploratory tool for interpreting complex milk quality data sets useful and effective for milk quality management.

5. REFERENCES

[1] Jenness, R., Wong, N.P., Marth, E.H. & Keeney, M. 1999. Composition of Milk. In: Fundamentals of Dairy Chemistry 3rd ed. Aspen Publishers Inc. Gaithersburg, Maryland, USA, pp: 1. [14] AOAC, 2000. Official Methods of Analysis International. 17th Ed. Association of Official Analytical Chemists, Washington. DC.

[2] Shah, S. K. & Khan, S. A., 1982. Factors influencing the protein level in

[4] Herrero-Martinez, J. M., Simo-Alfonso , E. F., Ramisramos, G., Gelfi, C., milk. Progressive Farming, 2(6):10-13.

[3] Fresno, J. M., Prieto, B., Urdiales, R., Sarmiento, R. M. & Carballo, J., 1995. Mineral content of some Spanish cheese varieties. Differentiation by source of milk and by variety from their content of main and trace elements. J. Sci. Food. Agr., 69: 339- 345. Righetti, P., Martinez, J.E.A. & Ramos, G., 2000. Determination of cow’s milk and ripening time in nonbovine cheese by capillary electrophoresis of the ethanol-water protein fraction. Electrophoresis, 21: 633-640.

[5] Martin-Hernandez, C., Amigo, L., Martin-Alvarez, P. & Juarez, M., 1992. Differentiation of milks and cheeses according to species based on the mineral content. Z. Lebensm. Unters. Forsch. 194: 541-544.

[6] Rodriguez, E. M. R., Alaejos, M. S., Rodriguez, E. M. R. R., Alaejos, M. S. & Romero, C. D., 1999. Chemometric studies of several minerals in milks. J. Agricult. Food. Chem., 47: 1520-1524.

[7] Favretto, L., Vojnovic, D. & Campisi, B., 1994. Chemometric studies on minor and trace elements in cow’s milk. Anal. Chim. Act., 293: 295- 300.

[8] Biagi, G., Valentini, A., Bagliacca, M., Corazza, M., Demi, S., Signorini, G. C., Greppi, G. F. & Romagnoli, A., 1990. Influenza del momento produttivo, dell’età e della stagione sul quadro lipidiconella capra Saanen. Ann. Fac. Med. Vet. Pisa, Italy, 43: 57-67.

[9] Biagi, G., Valentini, A., Bagliacca, M., Greppi, G. F., Signorini, G. C., Nannipieri, S., Romagnoli, A., 1991. Il quadro proteico nella capra Saanen: influenza dello stato fisiologico, dell’età e della stagione. Proc. 1st Congr. FeMeSPRum, Alghero, Italy, 1: 331-335.

[10] Harwood, E. D., Jensen, E. L., Wieckert, D. A. & Clayton, M., 1991. Milk yield variation concurrent with conception. J. Dairy Sci. 74: 2172-2179.

[11] Lattin, J., Carroll, D. & Green, P., 2003. Analyzing multivariate data. New York: Duxbury.

[12] Wunderlin, D. A., Diaz, M. D. P., Ame, M. V., Pesce, S. F., Hued, A. C. & Bistoni, M. D., 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River Basin (Cordoba Argentina). Water Res., 35: 2881–2894.

[13] Sanchez Lopez, F. J., Gil García, M. D., Martinez Vidal, J. L., Aguilera, P. A. & Garrido Frenich, A. 2004. Assessment of metal contamination in Doñana National park (Spain) using Crayfish (PROCAMBU RUS CLARKII). Environ. Moni. Assess., 93: 17–29.

[14] AOAC, 2000. Official Methods of Analysis International. 17th Ed. Association of Official Analytical Chemists, Washington. DC.

[15] Davide, C.L., 1977. Laboratory guide in dairy chemistry practicals. FAO Regional Dairy Development Centre for Asia and the Pacific. Dairy Training and Research Institute, Univ. Philippines, Los Banos, Laguna.

[16] Triebold, H.O., 2000. Quantitative Analysis with Applications to Agricultural and Food Products, Chapter XII, Second Printing, D. van Nostrand Company, Inc., New York, pp: 204-221.

[17] Perveen, F., Mumtaz, M. & Usmani, T.H., 2005. Estimation of metal contents in different varieties of milk available in Karachi city. J Chem Soc Pak., 27(6):611–614.