The Nucleus, **45** (3-4) 2008 : 137-141

The Nucleus
A Quarterly Scientific Journal of Pakistan
Atomic Energy Commission

NCLEAM, ISSN 0029-5698

APPLICATION OF PCA FOR THE CLASSIFICATION OF HUMAN WHOLE BLOOD SAMPLES

*S. RAHMAN and S. WAHEED

Chemistry Division, PINSTECH, P.O Nilore, Islamabad, Pakistan

(Received October 17, 2008 and accepted in revised form December 5, 2008)

Principal component analysis (PCA) is applied as a powerful tool to identify the possible correlations of different elements and patterns in large collection of blood samples. PCA has been successfully applied to analytical data of Cu, Cd, Li, Mg, Pb and Zn for their possible correlations with each other in 500 blood samples of healthy human subjects. The graphical representation of scores has been used to conceive the relative disparity in large collection of blood samples, while the loadings have been used to explain their any possible relationship among elements. Loadings show a direct correlation between Pb and Cd and negative correlation between Cu and Zn in the blood samples.The scores suggest 12 samples as outliers. These samples were rejected from the normal blood samples. However, the scores of the rejected blood samples indicate no defined correlation between Pb and Cd and inverse correlation between Cu –Zn, Mg -Cd and Cd - Li.

Keywords: Blood**,** Correlation coefficient, Principal component analysis, Elements

1. Introduction

For the formulation of baseline levels of different elements in blood a large number of blood volunteers from the normal population is mandatory that would generate analytical data sets of appropriate dimensions. The use of contemporary analytical techniques for the analysis of these samples is capable of producing massive amount of elemental data [1]. Sometimes the information gathered from such data may radically pose serious concerns to the human potential due to its enormity and complexity. Therefore, to extract valuable information from these voluminous and intricate findings, numerous mathematical approaches of data analysis are used to simplify the task of examining such results [2, 3]. The complex mathematical derivations and calculation related to these data analysis approaches have been further simplified in consequence to the use of modern computers using Excel spreadsheets. Chemometrics, in particular the principal components analysis (PCA) is probably the most widespread multivariate technique that can categorize the potential patterns in these tables [2, 4].

Since no authentic baseline values for the inorganic elements in human blood for the Rawalpindi and Islamabad region of Pakistan was documented, therefore it was the need of the times to plan, practically execute and develop a database for this issue. Moreover, the blood samples needs to be cautiously collected and interpreted when formulating its baseline elemental values, since these levels will in future be utilized to extrapolate for the deficiency or excess of these elements as a probable reason for the basis of different diseases. The literature shows the first use of PCA by Cauchy in 1829 but the specific references on its practical applications can be cited over the past few decades [5]. In view of the fact that no specific reference was available in published literature related to use of PCA for the evaluation of blood matrix, therefore the current paper attempts to apply PCA to a large collection of whole blood samples and the studied elements.

2. Experimental

2.1. *Sampling*

As a prerequisite for the sampling, detailed medical history of the normal healthy blood donors was recorded in the questionnaires designed for this study. Any anomalous finding regarding the subject’s socio-economic background, habit and health was also recorded. The blood samples were collected from the Federal Government Services

* Corresponding author : sohaila@pinstech.org.pk

Hospital and Shifa International Hospital, Islamabad under the supervision of trained medical staff by vein puncture and under contamination controlled conditions [6]. Minimum time was taken for transport of samples to the central laboratory and whenever required the samples were refrigerated at – 4 °C prior to analysis.

2.2. Sample preparation and equipment

About 0.5 mL of blood samples were taken in triplicate in 100 mL digestion flasks fitted with 30 cm long air condenser, 5mL distilled HNO_3 was added to the sample. The contents were heated at 80 °C for 30 minutes. After cooling 1.5 mL of concentrated perchloric acid (70%) was added and heated again at 250 ° C with occasional shaking till white fumes evolved. The clear solution obtained was cooled and transferred into a 10 mL measuring flask and the volume was made up with de-ionized water, for subsequent measurements of metals. A blank was prepared under similar conditions. Reference materials IAEA- H-8 (Horse Kidney) and IAEA -A-13 (Animal Blood) were also digested using the same procedure.

Atomic absorption spectrometry (AAS) and instrumental neutron activation analysis technique (INAA) were used to collect the elemental data. For AAS all the measurements were made with Hitachi model Z-180/80 polarized Zeeman atomic absorption spectrophotometer, which was coupled with a microprocessor-based data-handling facility and a printer. Hitachi model neon filled single element hollow cathode lamps of Cu, Zn, Mg, Pb and Cd were used as radiation sources. Concentration mode of flame atomic absorption spectrometer (FAAS) was used for the measurements of Cu, Zn and Mg [7], while the emission mode of FAAS was used for the measurements of Li. Electro thermal atomic absorption spectrometer (ETAAS) was used for the analysis of Pb and Cd [8].

For the INAA, one mL of whole blood was weighed and taken in a Vacuttee and was pre frozen (at 30 °C or lower) and subsequently freeze-dried in a BETA-A (Christ) freeze dryer for 48 - 72 hours or even more till complete dryness of the samples. After freeze drying, samples were homogenized in Fritsch Pulverizetts centrifuge ball mill for 15 min. After freeze drying blood it was observed that an average of 1gm liquid blood yields about 205 mg of dried powder. This factor was applied to correct the metal concentration in liquid samples with freeze dried powdered samples.

A synthetic comparison standard base was prepared for Zn and Mg using spec-pure salts in appropriate amounts dissolved in supra-pure aqua-regia acid mixture and used as a comparator. Freeze dried blood samples, multi-element synthetic standard alongwith IAEA reference material IAEA - H-8 (Horse Kidney) and IAEA - A-13 Animal Blood as control materials were taken in approximately 200 mg each and packed separately in pre-cleaned polyethylene irradiation capsules. Multiple target batches were prepared for irradiations and then packed in different irradiation rabbits. Pakistan Research Reactor II (PARR-II), which is a 27 kW, Miniature Neutron Source Reactor (MNSR) with a thermal neutron flux of $1x10^{12}$ $n.cm^{-2}$ s^{-1} was used for the optimized irradiations of the freeze dried blood samples. Cu, Mg and Zn were quantified using HPGe detector (Canberra Model AL-30) connected to PC-based Intertechnique Multichannel Analyzer (MCA) measuring system for the evaluation of spectra [9].

A total of 6 medically important elements namely Cu, Zn, Li, Mg, Pb and Cd were analysed in 500 whole blood samples of healthy subjects. These elemental concentrations were subjected to different statistical treatment tests. The blood samples along with the elemental data were then subjected to data processing for ultimate applications of principal component analysis using Excel to look for patterns in the data.

3. Data Analysis

3.1. Principal component analysis (PCA)

PCA was applied on the data for finding any trend. The basic model for PCA is as follows,

$$X = T \cdot P + E$$

Where T is the scores matrix that contains information of the samples in the new data space along columns, P is the loadings matrix that contains the weights of original variables in determining the new projection axis and E is the matrix of residuals which depends on the number of components selected to describe the model.

The output of PCA produced eigenvalues, Scores and Loadings matrices. [6, 10, 11]. Almost 80% of the total variation of the results was explained by the first three eigenvalues and retained to describe the data [1]. The scores matrix has 500 rows each representing individual sample. Due to extensive size of the scores and loadings they have not been incorporated in the paper. PCA model is applied to the analytical data and the trends are graphically visualized in easily interpretable plots [2, 12].

4. Results and Discussion

The scores have been graphically explained through different plots. These plots give an idea about the general trends between the samples and the presence of any outlier that deviates the normal patterns. Representative Fig. 1 show scores plotted between PC1 and PC2. The plot shows a cluster of samples in the middle of the figure indicative of the fact that all the samples represent a similar elemental pattern for normal blood subjects. However, the figure shows for few samples (Sample N2, N5, N102, N113, N116, N213, N224, N227, N324, N333, N338 and N435) that lie outside the main clusters. This gives a clue that these samples should be taken as outliers. The possible reasons for this abnormal behavior could be due to their medical history, family relationships, location of residence, eating habits, community, family status, sampling parameters, sample dissolution or instrumental variations. Therefore the elemental levels of these samples were not included in the results.

Similarly the loadings for the elements (Cu, Zn,

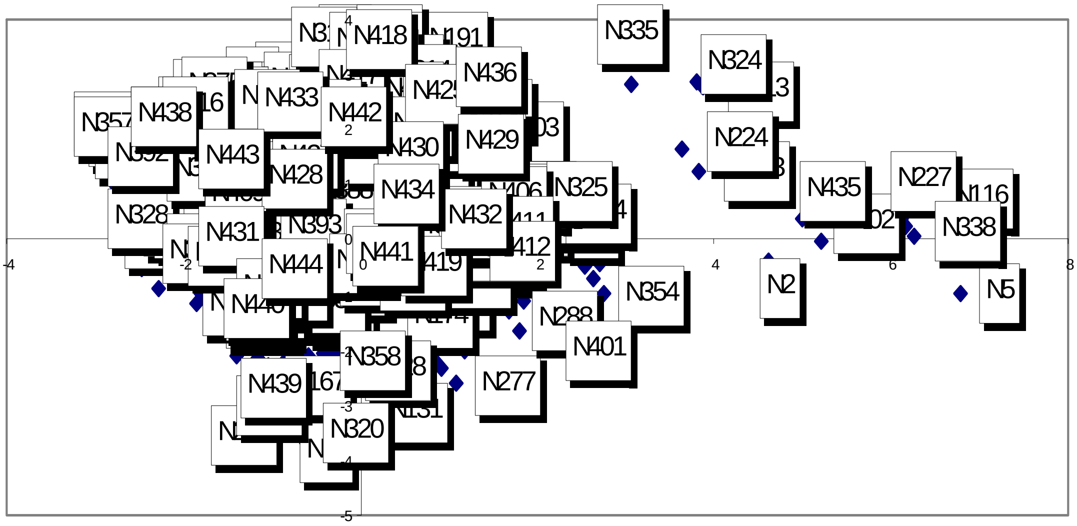

Figure 1. Scores of PC1 versus PC2 for blood samples of normal healthy subjects.

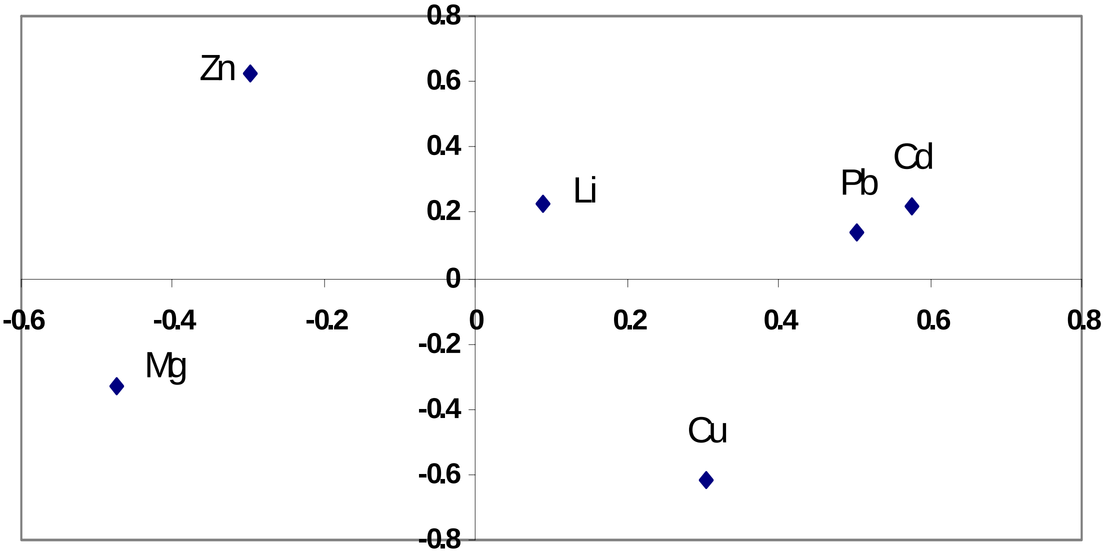

Figure 2. Loadings of PC1 versus PC2 for elements in blood samples of normal healthy subjects.

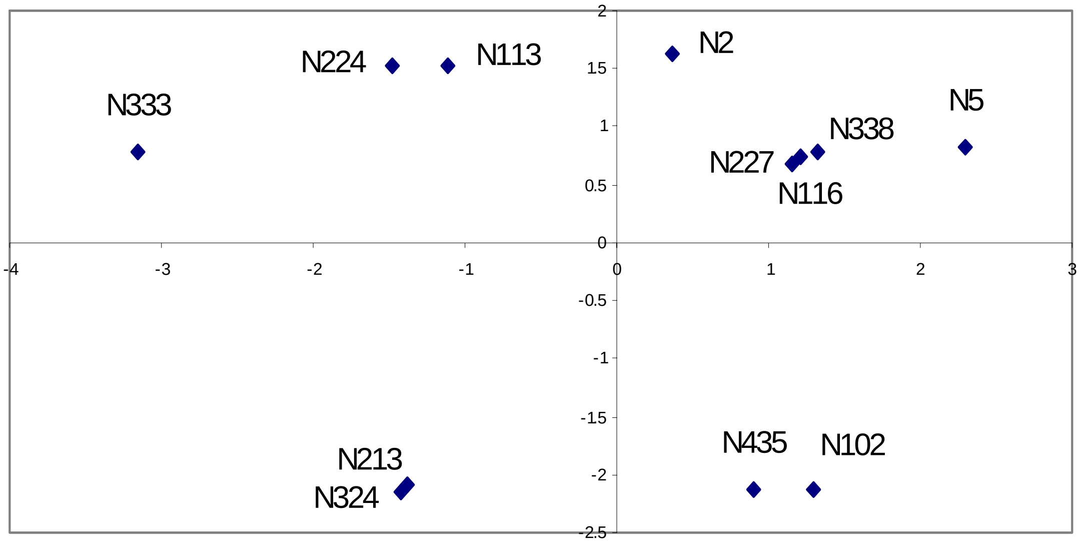

Figure 3. Scores PC1 Vs PC2 of the rejected normal blood samples.

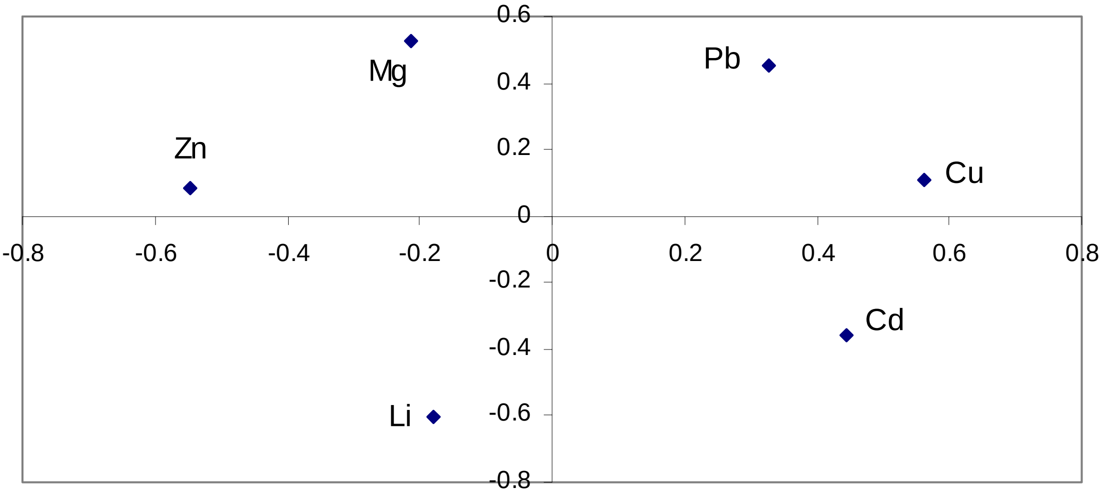

Figure 4. Loadings between PC1 and PC2 of the rejected blood samples.

Mg, Li, Pb and Cd) in blood are exploited graphically and represented as Fig. 2 between PC1 and PC2. The figure shows a correlation between Pd and Cd while Cu and Zn both lie showing a negative correlation.

4.1. Data analysis of the rejected blood samples

The scores plots for the analyzed blood of normal human subjects clearly predict that there are few samples out of 500 that deviate from the normal blood patterns. The data of these subjects were further scrutinized to look for any abnormal behavior in the analyzed elements or for other possible reasons. For this the standardized data of only these samples were subjected to PCA.

The scores from this PCA are represented as Fig. 3 as a plot between PC1 and PC2. The plot shows cluster of samples N116, N227 and N338. Similarly samples N224 and N113, N435 and N102, and N213 and N324 group together. However samples N2, N5 and N333 are very distinctly placed in the plot.

The correlations between the elements in these rejected samples were better perceived graphically as loadings in Fig. 4. No significant correlation between any of the studied elements is observed. There is no defined correlation between Pb - Cd as was evident in the case of rest of the blood samples. It shows a strong negative correlation between Pb - Li. Moreover. negative correlation between Cu – Zn and Mg – Cd is also evident. To comprehend for the possible reasons for the outlier blood samples, habit and health history of the donors were finally scrutinized. The blood samples (N116, N227, and N338) were found to have abnormally high Pb levels. This is probably due to their long-term low level exposure of Pb, as all three of these donors have their residence near industrial area of Islamabad. Two subjects (N213 and N324) were 56-57 years old and blood lead is generally higher for the older age groups. Out of these twelve subjects N2 and N5 were found to have very long family history of hypertension as a hereditary disease, although they did not develop any signs of high blood pressure probably due to their young age ranging between 25 to 27 years. One subject (N333) showed high Pb levels and belonged to very low socioeconomic class. The high Pb blood levels of this person could be due to the poor and improper locality where he lived under unhygienic conditions. Higher Pb and Cd were observed for two male subjects (N102 and N435). Both subjects were aged from 50-56. For remaining two samples (N113 and N224) no logical reasoning for their abnormally high Pb levels, were found therefore, in this case a possible contamination of the samples during different preparation stages is suggested.

5. Conclusions

The PCA applied to 500 blood samples of healthy donors gave scores and loadings matrices. The graphic representation of scores through different plots displayed 12 blood samples as outliers. The determined elemental concentrations of Cu, Cd, Li, Mg, Pb and Zn in these samples were not incorporated to establish the overall average baseline levels of healthy blood. Loading of the whole blood samples shows a positive correlation between Pb and Cd and negative correlation between Cu and Zn in the blood of normal subjects. However the scores of the rejected blood samples indicate no correlation between Pb and Cd and inverse correlation between Cu - Zn, Mg - Cd and Pb - Li. The low Pb and Cd correlation and the medical and socioeconomic parameters for these rejected subjects show abnormally high blood Pb levels.

Acknowledgements

We gratefully acknowledge all the normal healthy blood donors for their cooperation. The advice and the technical assistance of doctors and medical staff from Federal Government Services Hospital and Shifa International Hospital, Islamabad are also appreciated. Lastly all this would not have been possible without the software obtained from Professor Dr. R. G. Brerton, School of Chemistry, University of Bristol, UK.

References

[1] P. Van Espen and F. Adams, Analytica Chimica Acta **150** (1983) 153.

[2] B.G.M. Vandeginste, D.L. Massaert, L.M.C. Buydens, S. Dedjong, P.J. Lewi and J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part A&B, Elsevier, Amsterdam, (1998).

[3] E. J. Baum, Chemo and Intell Lab. Sys. **3** (1988) 91.

[4] S. Wold, C. Albano, W. J. Dunn, K. Esbensen, P. Geladi, S. Hellberg, E. Johanasson, W. Lindberg, M. Sjostrom, B. Skagerberg, C. Wikstrom and J. Ohman, Multivariate data analysis: Converting chemical data tables to plots", VII Internat. Conf. Computer Chem. Res. Educ. Garmisch-Partenkirchen, June 10-14, (1985) 166.

[5] R. G. Brerton, Chemometrics: Data analysis for the laboratory and chemical plant, Wiley, Chischester, (2003).

[6] D. Behne, J. Clin. Biochem. **19** (1981) 115.

[7] S. Rahman, N. Khalid, S. Ahmad, N. Ullah and M. Z. Iqbal, Pak. J. Med. Res. **43** (2004) 46.

[8] N. Khalid, S. Rahman, R. Ahmad and I. H. Qureshi, "Int. J. Environ. Anal. Chem. **28** (1987) 133.

[9] S. Waheed, J. H. Zaidi and S. Ahmad, J. Radioanal. Nucl. Chem. **258** (2003) 73.

[10] C. Chatfield and A. J. Collins, Introduction to multivariate analysis, Chapman and Hall, London, (1980).

[11] M. Wasim, M. S. Hassan and R. G. Brereton, Analyst **128** (2003) 1082.

[12] M. E. Chase, S. H. Jones, P. Hennigar, J. Sowles, G. C. H. Harding, K. Freeman, P. G. Wells, C. Krahforst, K. Coombs, R. Crawford, J. Pederson and D. Taylor, Mar. Pollut. Bull. **42** (2001) 491.