



ANÁLISE DE COMPONENTES PRINCIPAIS PARA SELEÇÃO DE VARIÁVEIS DE TESTES DE HOMOGENEIDADE E ESTABILIDADE EM ENSAIOS DE PROFICIÊNCIA



PRINCIPAL COMPONENT ANALYSIS FOR SELECTION OF VARIABLES IN HOMOGENEITY AND STABILITY TESTS APPLIED TO PROFICIENCY TESTING SCHEMES

ALBANO, Filipe^{1*}; CATEN, Carla ten¹; ANZANELLO, Michel¹

¹ Programa de Pós-Graduação em Engenharia de Produção (PPGEP) / UFRGS – Av. Osvaldo Aranha, 99 – Porto Alegre – Brasil.- Cep 90035-190 (fone: +55 (51) 3308-3336)

* Autor correspondente
e-mail: filipealbano@yahoo.com.br

Received 07 October 2016; received in revised form 31 October 2016; accepted 01 November 2016

RESUMO

Ensaio de Proficiência (EP) com base em comparações interlaboratoriais são atividades destinadas a avaliar a competência técnica de laboratórios na realização de medições específicas. As análises de homogeneidade e estabilidade das amostras preparadas são um item relevante no EP, no sentido de garantir a confiabilidade das rodadas de comparação, uma vez que a seleção inadequada de um parâmetro para realizar esta avaliação pode influenciar a comparação promovida. Esta pesquisa propõe um método para selecionar as variáveis mais relevantes em um EP, visando obter uma adequada representatividade nos testes de homogeneidade e estabilidade dos interlaboratoriais. A abordagem proposta se baseia em um índice de importância das variáveis analisadas em um EP, sendo obtida através da Análise de Componentes Principais (ACP). O método proposto foi aplicado em três EP brasileiros, na área de análise de vinho tinto, águas e carvão. Os resultados indicam que o uso de ACP foi adequado para ajudar a seleção de variáveis para realização dos testes de homogeneidade e estabilidade de programas interlaboratoriais. As variáveis selecionadas também foram analisadas criticamente por especialistas nas áreas dos EP, os quais confirmaram a adequação do método proposto.

Palavras-chave: Ensaio de proficiência, comparação interlaboratorial, homogeneidade, estabilidade, análise de componentes principais.

ABSTRACT

Proficiency Tests (PT) based on interlaboratory comparisons are activities aimed at assessing the technical competence of laboratories in carrying out specific measurements. The analyses of homogeneity and stability of prepared samples are an important step in ensuring the reliability of the comparison rounds, since improper selection of the parameter to carry out this evaluation can influence the promoted comparison. This paper proposes a method for selecting the most relevant variables aimed at improving homogeneity and stability tests in PT. For that matter, the approach relies on a variable importance index derived from Principal Components Analysis (PCA) parameters. The proposed method was applied to three different PT schemes (beverage, water and coal) in Brazil. Results indicate that the use of PCA was adequate to help the variable selection of homogeneity and stability tests in PT schemes. The selected subset of variables was corroborated by experts in the PT schemes analyzed.

Keywords: Proficiency tests, interlaboratory comparison, homogeneity, stability, principal component analysis.

INTRODUCTION

Laboratories seeking accreditation should participate in PT Schemes, which are important instruments that provide confidence for testing and calibration (ISO, 2010). In the testing area, one of the important items in PT is the preparation of samples, which must ensure that such samples are homogeneous and stable so that the variability between them is negligible across the variability of laboratories (ISO, 2015). That becomes a relevant fact, since the non-homogeneity or stability of a batch of samples can provide the laboratory with a false impression of adequate or inadequate results in a comparison program.

Homogeneity and stability tests should always be conducted in PT in order to ensure the reliability of the comparison rounds (ISO, 2015). These tests are often expensive, and it typically becomes hard to choose which chemical or biological variable, among others, should be used in the homogeneity and stability tests (THOLEN *et al.*, 2006). Since it is often economically unfeasible to test all variables, the most critical ones (higher breakdown or more sensitive) are chosen to conduct the tests. Despite its importance, studies attesting the correlation between variables and the real representativeness of the variable chosen by experts to represent the sample are rarely carried out (ALBANO; CATEN, 2014).

Baumeister (2013) inform that an assessment of homogeneity must be conducted with appropriate statistical designs. The results of homogeneity and stability tests must be analysed statistically and may be included in the final report (PTA, 2016). Also, every accredited PT scheme provider must be able to justify the claim that its samples are homogenous and stable (ILAC, 2008). The homogeneity and stability tests need to be fit for purpose, because there are many possibilities to develop a PT scheme in different areas (Thompson, 2016).

In light of the aforementioned, we present the following research questions: how to discover if the variable chosen by experts to test homogeneity and stability of the PT samples really represents the variability in sample preparation? Is selecting a single variable sufficient for conducting homogeneity and stability tests?

The Annex B of the ISO 13528 standard addresses the assessment of homogeneity and

stability tests. It takes at least 10 samples analyzed twice (each sample) to evaluate homogeneity and 2 samples analyzed in duplicate to assess stability. The tests shall be performed in repeatability conditions (ISO, 2015). A PT with many variables involved can require an elevated number of homogeneity and stability tests, which can become unaffordable in terms of cost and time demanded for analysis. The standards related to PT schemes do not present clear criteria on the systematics or methods for selecting the variables to be used in homogeneity and stability tests in PT (ALBANO; CATEN, 2014).

To examine the homogeneity and stability of samples it is necessary to select at least one variable representing the samples being compared. Any failure in this process can compromise a PT round. A reliable course of action tailored to select variables is the use of Principal Components Analysis (PCA), which allows identifying the variables that better represent the variability of data and increases the efficiency of their interpretation (HAIR, 2009).

For Owen (2014), the primary purpose of principal component analysis (PCA) is to reduce the dimension of a large data set containing interrelated variables into a more concise data set that retains most of the existing variations.

Mathematically, PCA performs an orthogonal transformation to convert a set of correlated variables in a set of new uncorrelated variables by linearly combining the original variables. This procedure yields the so called principal components (PC), which are typically in smaller number when compared to the original variables (RENCHE, 1995). According to Ringnér (2008), PCA is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set. It accomplishes this reduction by identifying directions, called principal components.

The transformation of data carried out by PCA aims at maximizing the variance between the PCs, and is performed in such a way that the first PC represents the largest possible variance, the second represents the second larger variance and so on (ROBERTS *et al.*, 2012). Besides reducing the problem dimensionality, the new variables generated by PCA are orthogonal between themselves, justifying the wide use of that technique in datasets affected by highly correlated variables (HAIR, 2009).

In terms of its practical contributions, PCA has been widely applied in several segments with different scopes: (i) exploratory data analysis, preparation of forecasting and variable selection models (ANZANELLO *et al*, 2009); (ii) analysis and discussion in terms of the components' scores, also called factor scores and their loadings (BRO; SMILDE, 2014); (iii) reduction of the number of original variables (REIS *et al*, 2010; PAUL *et al*, 2013); (iv) unveiling of internal structure of data aimed at obtaining proper explanation of its variance (ANZANELLO *et al*, 2009; ZOU *et al*, 2006); (v) graphical analysis and clustering of observations; (vi) PCA biplot used to examine the multi-dimensional variation and the relationships among the variables (ALKAN *et al*, 2015); (vii) uses of principal components control charts to explain variability and establish a method of monitoring control future observations (LAZAROTTO, 2016; MARTINS *et al*, 2012); (viii) possibility to work with missing data matrix in a PCA robust analyses (EMMANUEL *et al*, 2011); (ix) possibility of face recognition using PCA (PAUL; SUMMAN, 2012); and, in the last 40 years, (x) the extensive use of PCA to propose different methods to perform variable selection (KOWALSKI; BENDER, 1974).

This paper proposes a PCA-based method to select relevant variables in homogeneity and stability tests of samples in PT. For that matter, loadings generated by the PCA give rise to a variable importance index aimed at identifying the most important variables to be assessed in PT. Variables presenting the highest indices are qualitatively assessed by process experts, and a decision on their retention is made. As secondary objectives, the following aspects stand out: (i) application of the proposed framework to three different PT schemes; and (ii) discussion on the advantages and disadvantages of applying PCA in PT scenarios.

PT SCHEMES DATA

The developed research presents an applied character and follows a quantitative approach. As for the objectives, this is an exploratory work that uses as technical procedure a case study applied through the analysis of real data from PT schemes promoted by Rede Metrológica RS, a Brazilian accredited ISO/IEC 17034 provider.

The study was conducted in three different PT schemes (beverages – red wine,

water – waste water, and mineral coal), based on chemical analyses. Details on the data of PT are depicted in Table 1.

Table 1. *PT schemes data details*

The missing data in such datasets was replaced using an interpolation technique that relies on solving a direct linear system of equations for missing elements; that procedure was carried out in Matlab® 2014. The software used to perform the PCA analyses was SPSS (Statistics for Windows, version 21.0., Armonk, NY).

Results were obtained in a fourteen-day interval and samples were evaluated by different laboratories. The comparison was aimed at assessing the performance of the various participants and checking whether they were technically competent in measuring the proposed variables. The PT schemes were structured according to ISO/IEC 17043 and the performance was evaluated according to ISO 13528 standard (through the consensus value – Annex C).

RESULTS AND DISCUSSION

We initially assessed the correlation matrix of the original data of each PT, depicted in Tables 2, 3 and 4. Some variables are highly correlated, especially in mineral coal data (see Table 4).

Table 2. *Beverage variables correlation matrix*

Table 3. *Waste water variables correlation matrix*

Table 4. *Mineral coal variables correlation matrix*

PCA was then applied to the correlation matrix of each PT; parameters of interest emerging from PCA include the amount of variance explained by the j th retained PC, and the loading of variable i in each PC j , w_{ij} . For the variable selection procedure we propose an importance index, vs_i , to highlight variables that explain most variance on data; see Equation (1). The proposed index relies on the following rationale: variables with high squared loadings (w_{ij}^2) contribute substantially on explaining variability inside a specific PC. In our propositions, we retained a number of PCs that

explains at least 70% of the cumulative variance in the data. Next, we assess the variable responsible for the highest vs on each retained PC, starting from PC1 until the last retained PC. Such variables are then qualitatively assessed by process experts for a final decision on retaining or not that variable for PT.

$$vs_{ij} = w_{ij}^2 \text{ for } ij = 1, \dots, l \quad (1)$$

Tables 5, 6 and 7 depict the PCA loadings and estimated vs for the three PT schemes (beverage, waste water and mineral coal). Once the recommended variables are selected by the proposed index, experts qualitatively assess whether such variables are suitable for PT purposes.

As for the PT of Beverage analyses (Red wine), the variables with highest vs were 1-propanolol (PC1), pH (PC2), Sugar (PC3) and chloride (PC4). In Table 5, vs of such variables are highlighted. Those variables were qualitatively assessed by a process expert with more than 20 years of experience in wine analyses, who also is the technical responsible for one of the official beverage laboratories of Brazilian government. According to his opinion, parameters 1-propanolol, pH and sugar properly represent the wine sample: (i) 1-propanolol, which represents the group of chromatographically testing (such variable is moderately correlated ($|r| > 0.6$) with ethyl acetate, acetaldehyde, 2-methyl propanol and 3-Methyl-2-butanol); (ii) pH represents the group of traditional physic-chemical analyses and is based on a method that relies on a direct measurement from the sample; and (iii) sugar, which represents the volumetric analysis (titrations) group. In addition, the expert also commented that these three variables (traditional physic-chemical and chromatographical) have different magnitudes in terms of uncertainty measurement. The selected variables also are sensible in terms of stability, what is relevant for the PT purpose. Finally, the expert added that chloride should not be considered relevant at this time as it can be represented by other variables, and analysed by the same kind of method (chromatographical, titrimetric or potentiometric).

Table 5. *Principal components of Beverage PT*

Table 6 depicts the results of vs for the waste water PT scheme. The expert, who presents 26 years of experience in water analysis in an accredited ISO / IEC 17025 laboratory, reported that the Pb parameter (responsible for the highest vs in CP1) is a suitable parameter for the homogeneity and stability testing in PT, since it properly represents the group of metals. In addition, the expert complemented his assessment saying that the Ca (responsible for the highest vs in CP2) should also be considered as it can be analysed by different methods and represents another group of metals. The variables with the highest vs in CP3 and CP4 (Sr and As) do not include relevant information when compared to that captured by Pb and Ca. Based on such assessment, we understand that Pb and Ca can properly represent the group and are suitable for ensuring homogeneity and stability test in this PT.

Table 6. *Principal components of Waste Water PT*

Results of the variable selection procedure for the Mineral Coal PT scheme are presented in Table 7. The expert consulted is a manager of a fuel laboratory accredited ISO/IEC 17025 with more than 30 years of experience in coal analysis. The highest vs values in CP1, accounting for 84,7% of the variance, were provided by total sulfur (Total_S), volatile matter (Volat_mat) and fixed carbon (Fixed_Carbon). Moisture also yielded a high vs , but such variables was not considered as representative since all results were reported on a dry basis. The sulfur and volatile matter are the most sensitive variables due to potential oxidation, while fixed carbon is directly related to volatile matter calculation (justifying its high vs score). On the other hand, ash is a mineral residue and presents very stable properties, so it is not representative in a PT.

The expert recommend using variables total sulfur or volatile matter in the analysis. In addition, the volatile matter is also recommended in the homogeneity and stability tests, since it is cheaper to be obtained. In cases like this one, PCA can gives little guidance since all variables, except Ashes, are very highly correlated. Based on this comment, variable selection should proceed on the basis of expert evaluation of the physicochemical (and analysis cost) properties of the measurements.

Table 7. Principal components of Mineral Coal PT

In order to develop this research, it is necessary to rely on real PT databases, whereas it is not possible to carry out the proposed steps without prior rounds. However, it is not always possible to count on databases comprised of all tests; using algorithms to fulfill the missing data may affect the actual representation of variability retained by the PC, requiring deeper analysis.

In addition, Monte Carlo techniques could also be considered as alternative ways for estimating uncertainty estimates on the loadings, especially when considering the problem of missing data (SUNG; GEYER, 2007). This approach is seen promising in future developments of the framework here proposed.

CONCLUSIONS

Selection of representative variables from one sample for homogeneity and stability testing is a significant event in the planning and execution of a PT. Such variables should be able to represent the behavior of the sample during the period of PT implementation. The application of PCA, as presented this paper, was deemed helpful to define the most representative variables on a PT. For that matter, parameters emerging from the PCA were used to build a variable importance index aimed at identifying the most important variables to be evaluated in PT. Variables yielding the highest indices were then qualitatively assessed by process experts regarding their retention or removal in PT.

When applied to data from three Brazilian PT programs, the proposed framework led to satisfactory results as it was able to consistently select variables to evaluate the homogeneity and stability of the PT samples in beverage, waste water and mineral coal datasets. The variables pointed as relevant by the method were corroborated as important when qualitatively assessed by process experts.

Future developments include the application of the method proposed in other PT schemes in different areas, and the proposition of homogeneity and stability tests based on the results of the participating labs.

REFERÊNCIAS:

1. ALBANO, F, CATEN, C (2014). Proficiency tests for laboratories: a systematic review. **Accred Qual Assur** 19:245–257. doi 10.1007/s00769-014-1061-8
2. ANZANELLO, M. ALBIN, S. CHAOVALITWONGSE, W. Selecting the best variables for classifying production batches into two quality levels. **Chemometrics and Intelligent Laboratory Systems**. v. 97, 2009. <http://dx.doi.org/10.1016/j.chemolab.2009.03.004>.
3. B. Baris Alkan , Cemal Atakan, Yilmaz Akdi (2015). Visual Analysis Using Biplot Techniques of Rainfall Changes over Turkey. **Mapan Journal**. Vol 30, Issue 1 , pp 25-30. Doi 10.1007/s12647-014-0119-8
4. BAUMEISTER, F. Homogeneity of EQA samples – requirements according to ISO/IEC 17043. **EQALM Symposium**, Bucharest. 2013.
5. Bro, R. Smilde, A (2014). Principal component analysis. **Anal. Methods - The Royal Society of Chemistry**. Vol 6, 2812–2831 . Doi: 10.1039/c3ay41907j
6. EMMANUEL, J. CANDE, S. LI, X. MA, Y. WRIGHT, J (2011). Robust Principal Component Analysis? **Journal of the ACM**, Vol. 58, No. 3, Article 11. May 2011.
7. HAIR JR, ANDERSON RE, TATHAM RL & BLACK WC. **Multivariate Data Analysis** (7th Edition). 2009. ISBN-13: 978-0138132637
8. ILAC. **Discussion Paper on Homogeneity and Stability Testing**. ILAC Proficiency Testing Consultative Group, Task Force on H&S testing. 2008.
9. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 13528** (2015) Statistical methods for use in proficiency testing by interlaboratory comparisons. Geneva.
10. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 17043** (2010). Conformity assessment - General requirements for proficiency testing. Geneva.
11. KOWALSKI, B. BENDER, C (1974). An orthogonal feature selection method. **Pattern Recognition** 8(1):1-4. Doi: 10.1016/0031-3203(76)90023-6.

12. LAZZAROTTO, E (2016). Principal components in multivariate control charts applied to data instrumentation of dams. *Independent Journal of Management & Production (IJM&P)*. DOI: 10.14807/ijmp.v7i1.369
13. MARTINS, S. GODOY, L. LOPES, L (2012). Monitoramento do controle estatístico do processo utilizando ferramentas estatísticas. *Revista Espacios*. Vol 33 (Nº 6).
14. OWEN, J (2014). Principal Component Analysis: Data Reduction and Simplification. *McNair Scholars Research Journal*.
15. PAUL, L. SUMAM, A (2012). Face Recognition Using Principal Component Analysis Method. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*. ISSN: 2278 – 1323. Volume 1, Issue 9.
16. PAUL, L; SUMAN, A. SULTAN, N. Methodological Analysis of Principal Component Analysis (PCA) Method (2013). *IJCEM International Journal of Computational Engineering & Management*, Vol. 16 Issue 2, ISSN (Online): 2230-7893.
17. PROFICIENCY TESTING AUSTRALIA. **Guide to proficiency testing Australia**. PTPM 1.1.06 May, 2016.
18. REIS, P. *ET AL*. Mineral composition of Lippia alba (Mill.) N.E. Brown leaves. *J. Braz. Chem. Soc.* vol.21 no.10 São Paulo 2010.
- <http://dx.doi.org/10.1590/S0103-50532010001000015>
19. RENCHER, A. C. **Methods of Multivariate Analysis**. New York: Wiley, 1995.
20. RINGNÉR, M (2008). What is principal component analysis? *Nature Biotechnology* 26, 303–304. DOI:10.1038/nbt0308303
21. ROBERTS Y. L. *et al.* Quantitative comparison of the variability in observed and simulated shortwave reflectance. In: *Atmos. Chem. Phys. Discuss.* 12, 28305-283041, 2012. Doi:10.5194/acpd-12-28305-2012.
22. SUNG, Y. GEYER, C (2007). Monte carlo likelihood inference for missing Data models. *Institute of Mathematical Statistics*. Vol. 35, No. 3, 990–1011. Doi: 10.1214/009053606000001389
23. THOLEN, D. CHAPPEL, R. FRANCIS, B. *Accred Qual Assur.* 11: 400–407 DOI 10.1007/s00769-006-0167-z. 2006.
24. THOMPSON, M. *Statistics and Data Analysis in Proficiency Testing*. University of London. Available at: <http://www.rsc.org/images/madrid%202007_tcm18-87000.pdf> Accessed in October, 2016.
25. ZOU, H. HASTIE, T. TIBSHIRANI, R (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, Volume 15, Number 2, Pages 265–286. DOI: 10.1198/106186006X113430

Table 1. PT schemes data details

PT scheme	Variables	Laboratories (N)	Variables (N)
Beverage (Red Wine)	1-propanol (mg/100mL), ethyl acetate (mg/100mL), acetaldehyde (mg/100mL), 2-methyl propanol (mg/100mL), ashes (g/L), sugar (g/L), pH (-log[H ⁺]), chlorides (mg/L), malic acid, alcohol (%), density (g/cm ³), total acidity (mEq/L), free SO ₂ (mg/L), total SO ₂ (mg/L), volatile acidity (mEq/L), adjusted volatile acidity (mEq/L), methanol, 3-Methyl-2-butanol (mg/100mL)	32	18
Waste Water	As, Fe, Ni, Hg, Ca, Cd, Pb, Cr, Na, Ba, Mn, Sr (mg/L)	77	12
Mineral Coal	moisture (%), ashes (%), volatile matter (%), fixed carbon (%), total S (%), calorific power (J/g)	20	6

Table 2. Beverage variables correlation matrix

Correlation	sugar	Alc.	Dens.	Vol. Acid.	Ajus. Vol. Acid.	Tot. Acid.	pH	free_S O ₂	total_SO ₂	Meth.	Acet.	Eth. Acet.	Malic Acid	1-prop.	2-met prop	3-met 2-but	ashes	Chlo.
sugar	1.00	-0.28	0.08	0.00	-0.55	0.08	-0.05	-0.43	0.14	-0.66	0.05	0.18	0.33	0.12	0.20	-0.05	0.15	0.04
Alc.		1.00	-0.05	0.25	0.30	0.63	-0.71	0.31	0.44	-0.01	0.24	0.01	-0.22	-0.12	-0.01	0.20	-0.10	0.13
Dens.			1.00	-0.10	-0.09	-0.06	0.04	-0.12	-0.07	-0.18	0.03	-0.10	-0.16	-0.09	-0.18	-0.18	-0.07	-0.12
Vol. Acid.				1.00	0.36	0.53	-0.48	-0.09	0.33	0.07	-0.25	-0.19	0.05	-0.34	-0.24	0.41	0.15	0.25
Ajus.Vol. Acid.					1.00	0.51	-0.51	0.52	0.25	0.49	-0.13	-0.27	-0.23	-0.25	-0.13	0.27	-0.06	0.02
Tot. Acid.						1.00	-0.94	0.18	0.63	0.03	0.05	-0.13	-0.19	-0.27	-0.07	0.30	-0.13	0.09
pH							1.00	-0.22	-0.66	-0.01	-0.09	0.13	0.17	0.24	0.03	-0.30	0.06	-0.12
free_SO ₂								1.00	0.02	0.35	0.19	0.00	-0.19	0.04	0.12	0.11	-0.18	-0.08
total_SO ₂									1.00	0.00	-0.24	-0.33	-0.21	-0.38	-0.11	0.43	-0.14	0.01
Meth.										1.00	-0.16	-0.23	-0.11	-0.15	-0.06	0.37	-0.12	0.01
Acet.											1.00	0.75	0.09	0.70	0.40	-0.66	0.11	-0.07
Eth. Acet.												1.00	0.45	0.85	0.64	-0.61	0.35	0.08
MalicAcid													1.00	0.53	0.55	-0.13	0.28	0.29
1-prop.														1.00	0.80	-0.68	0.49	0.22
2-met prop															1.00	-0.21	0.53	0.33
3-met 2-but																1.00	-0.07	0.21
ashes																	1.00	0.68
Chlo.																		1.00

Table 3. Waste water variables correlation matrix

Correlation	As	Fe	Ni	Hg	Ca	Cd	Pb	Cr	Na	Ba	Mn	Sr
As	1.00	0.11	-0.01	-0.01	-0.10	0.14	0.12	-0.03	-0.17	-0.01	0.23	-0.04
Fe		1.00	0.47	0.43	0.02	0.40	0.43	0.42	-0.15	0.32	0.47	0.09
Ni			1.00	0.16	-0.07	0.76	0.77	0.77	-0.45	0.52	0.42	-0.16
Hg				1.00	0.19	0.09	0.11	0.12	0.12	0.10	0.26	0.14
Ca					1.00	-0.17	-0.12	-0.26	0.43	-0.19	-0.22	-0.32
Cd						1.00	0.87	0.68	-0.55	0.48	0.52	-0.13
Pb							1.00	0.74	-0.57	0.61	0.60	-0.19
Cr								1.00	-0.36	0.65	0.48	-0.16
Na									1.00	-0.43	-0.41	-0.08
Ba										1.00	0.45	0.00
Mn											1.00	-0.01
Sr												1.00

Table 4. Mineral coal variables correlation matrix

Correlation	moisture	ashes	volatile mat.	fixed carbon	total S	calorific power
moisture	1.00	0.28	-1.00	-1.00	-1.00	0.99
ashes		1.00	-0.27	-0.29	-0.27	0.32
volatile matter			1.00	1.00	1.00	-0.99
fixed carbon				1.00	1.00	-0.99
total S					1.00	-0.99
calorific power						1.00

Table 5. Principal components of Beverage PT

Variable	PC and squared loadings (vs)			
	1	2	3	4
Sugar	0.10	0.00	0.57	0.11
Alc.	0.20	0.37	0.02	0.08
Dens.	0.00	0.05	0.00	0.14
Vol. Acid.	0.25	0.11	0.14	0.04
Ajus.Vol. Acid.	0.37	0.11	0.18	0.05
Tot. Acid.	0.39	0.40	0.02	0.08
pH	0.37	0.47	0.02	0.08
free_SO2	0.06	0.08	0.46	0.00
total_SO2	0.37	0.11	0.10	0.04
Meth.	0.12	0.00	0.30	0.30
Acet.	0.27	0.24	0.15	0.20
Eth. Acet.	0.56	0.23	0.02	0.02
MalicAcid	0.27	0.07	0.08	0.10
1-prop.	0.71	0.21	0.04	0.00
2-met prop	0.35	0.33	0.00	0.06
3-met 2-but	0.49	0.00	0.08	0.19
Ashes	0.16	0.17	0.09	0.24
Chlo.	0.01	0.20	0.12	0.33
Eigenvalue	5.07	3.16	2.40	2.07
% of Variance	28.16	17.57	13.32	11.48
Cumulative %	28.16	45.73	59.05	70.53

Table 6. Principal components of Waste Water PT

Variable	PC and squared loadings (vs)			
	1	2	3	4
As	0.02	0.03	0.02	0.82
Fe	0.33	0.18	0.17	0.00
Ni	0.72	0.02	0.03	0.02
Hg	0.04	0.42	0.30	0.00
Ca	0.07	0.57	0.09	0.01
Cd	0.76	0.00	0.02	0.00
Pb	0.84	0.00	0.03	0.00
Cr	0.71	0.00	0.02	0.04
Na	0.40	0.27	0.00	0.00
Ba	0.51	0.00	0.00	0.05
Mn	0.50	0.00	0.08	0.06
Sr	0.01	0.10	0.62	0.08
Total (Eigenvalue)	4.91	1.59	1.39	1.11
% of Variance	40.94	13.29	11.62	9.24
Cumulative %	40.94	54.23	65.84	75.09

Table 7. Principal components of Mineral Coal PT

Variable	PC and squared loadings (vs)	
	1	2
Moisture	0.99	0.01
Ashes	0.12	0.87
Volat_mat	0.99	0.01
Fixed_Carbon	0.99	0.01
Total_S	0.99	0.01
calorific_power	0.98	0.01
Eigenvalue	5.082	0.901
% of Variance	84.703	15.009
Cumulative %	84.703	99.712