Proceedings of the International Conference on Manufacturing Systems – ICMaS
Vol. 4, 2009, ISSN 1842-3183
University POLITEHNICA of Bucharest, Machine and Manufacturing Systems Department
Bucharest, Romania

COMPUTER AIDED SELECTION IN DESIGN PROCESSES WITH MULTIVARIATE STATISTICS

Adrian PARIS, Constantin TÂRCOLEA

Abstract: : *Principal Component Analysis (known briefly as PCA) is a multivariate statistical technique for simplifying a cloud data set [1, 2]. Based on this procedure the observable possibly correlated properties are reported into a few uncorrelated "attributes"; in other words it is considered a transformation from a space into a subspace such that the retained variance of the original cloud is "maximal" by this new representation [3, 4]. In a previous paper [7] we applied the so called the Jöreskog' technique used for the dimensional reduction in a bivariate subspace. The goal of this paper is to apply the Pearson method of PCA, for the same family of materials, using the XLSTAT 2009 software; new relevant factors are obtained, and the results are comparable.*

Key words: *principal component analysis, materials selection, artificial factors, design of experiments.*

1. INTRODUCTION

The Pearson's method minimizes redundancy, measured by correlations, and maximizes the signal, measures by variance. The procedure is based in part on an iterative algorithm. The technique begins by finding a normalized direction along which the variance is maximized. Next it is find the second direction along which the remains variance is maximized, under restricting the search to all directions perpendicular to previous selected directions. This iterative procedure is repeated until it is obtained all vectors.

2. AN APPLICATION OF PCA BASED ON PEARSON'S TECHNIQUE

Principal component analysis is a standard technique to reduce multivariate data sets to lower dimensions [5]. The number of observable attributes gives the dimension of the initial vector space of the objects. The PCA model represents the objects in view in a strictly subspace [6].

Instead of real attributes, the PCA proposes new factors, but artificial ones, so that the subspace yields the minimum deformation of the original cloud. In the present paper the dimensional reduction for a family of materials (Table 1), uses the Pearson's method [7, 8].

Let us consider X as a 6×9 dimensional matrix, attributes/materials constructed for the first six attributes from Table 1 [7]:

$$X = \begin{pmatrix} 1 & 2 & 3 & 3 & 3 & 2.5 & 3.5 & 4 & 4 \\ 1 & 2 & 3 & 3 & 3 & 2.5 & 3 & 4 & 5 \\ 1 & 2 & 2 & 3 & 2 & 2 & 1 & 1 & 1 \\ 4 & 3 & 2 & 2 & 2 & 2 & 3 & 4 & 5 \\ 4 & 4 & 3 & 3 & 3 & 2 & 4 & 5 & 5 \\ 5 & 3 & 4 & 3 & 2 & 2 & 3.5 & 3 & 2 \end{pmatrix}.$$

The standardized matrix associated with X is:

$$Z = \begin{pmatrix} -1.96 & -0.92 & 0.11 & 0.11 & 0.11 & -0.40 & 0.63 & 1.15 & 1.15 \\ -1.72 & -0.83 & 0.05 & 0.05 & 0.05 & -0.39 & 0.05 & 0.93 & 1.82 \\ -0.94 & 0.47 & 0.47 & 1.88 & 0.47 & 0.47 & -0.94 & -0.94 & -0.94 \\ 0.89 & 0 & -0.89 & -0.89 & -0.89 & -0.89 & 0 & 0.89 & 1.79 \\ 0.33 & 0.33 & -0.66 & -0.66 & -1.66 & 0.33 & 1.33 & 1.33 & 1.33 \\ 1.92 & -0.05 & 0.93 & -0.05 & -1.04 & -1.04 & 0.44 & -0.05 & -1.04 \end{pmatrix}.$$

Table 1

Properties and materials for sliding bearing [7]

Materials	**Alloys on the basis of**	**Bronze on the basis of**		**Aluminum Alloys**	**Porous sintered bearings**			**Plastics**	**Artificial carbon**
Properties	Lead	Tin	Lead	Tin	Al				
Sliding properties	1	2	3	3	3	2.5	3.5	4	4
Embeddability	1	2	3	3	3	2.5	3	4	5
Emergency running (antifrictionnal) properties	1	2	2	3	2	2	1	1	1
Loadability	4	3	2	2	2	2	3	4	5
Heat conduction /thermal expansion	4	4	3	3	3	2	4	5	5
Corrosion resistance	5	3	4	3	2	2	3.5	3	2
Minimal or dry lubrication	2	3	4	5	4	3	1	1	1

Table 2

Summary statistics Principal Component Analysis (PCA) (Tables 3–8 and Figs. 1, 3 and 4)

XLSTAT 2009.2.03 – Principal Component Analysis (PCA) PCA type: Pearson (n)
Type of biplot: Correlation biplot / Coefficient = Automatic Summary statistics:

Var	Obs	Obs. with missing data	Obs. without missing data	Min.	Max	Mean	Std. Dev.
Var1	9	0	9	1.000	4.000	2.889	0.961
Var2	9	0	9	1.000	5.000	2.944	1.130
Var3	9	0	9	1.000	3.000	1.667	0.707
Var4	9	0	9	2.000	5.000	3.000	1.118
Var5	9	0	9	2.000	5.000	3.667	1.000
Var6	9	0	9	2.000	5.000	3.056	1.014

Using the matrix equation

$$R = \frac{1}{8} Z * Z_{tr}, \quad (1)$$

it results that:

$$R = \begin{pmatrix} 1 & 0.94 & -0.15 & 0.17 & 0.34 & -0.54 \\ 0.94 & 1 & -0.18 & 0.34 & 0.42 & -0.59 \\ -0.15 & -0.18 & 1 & -0.79 & 0.71 & -0.23 \\ 0.17 & 0.34 & -0.79 & 1 & 0.89 & 0.11 \\ 0.34 & 0.42 & 0.71 & 0.89 & 1 & 0.14 \\ -0.54 & -0.59 & -0.23 & 0.11 & 0.14 & 1 \end{pmatrix},$$

which is the correlation matrix. The proximity between attributes is expressed in terms of correlations.

XLSTAT is a Microsoft Excel, add-in that has been developed since 1993, to enhance the analytical capabilities of Excel. XLSTAT relies on Excel for the input of data and the display of results, but the computations are done using autonomous software components. Main product of the company Addinsoft (`www.xlstat.com`) from the applications XLSTAT group is XLSTAT-Pro, with the most important features: PREPARING DATA; DESCRIBING DATA (with the PCA application); ANALYZING DATA; VISUALIZING DATA; MODELING DATA AND FORECASTING; CORRELATION AND ASSOCIATION TESTS; PARAMETRIC TESTS; NON PARAMETRIC TESTS.

We applied the version 2.03 of this software and the obtained results are recorded bellow:

Table 3

Correlation matrix

Correlation matrix (Pearson(n)):						
Var.	Var1	Var2	Var3	Var4	Var5	Var6
Var1	**1**	0.943	-0.153	0.175	0.347	-0.538
Var2	0.943	**1**	-0.182	0.346	0.424	-0.597
Var3	-0.153	-0.182	**1**	-0.791	-0.707	-0.232
Var4	0.175	0.346	-0.791	**1**	0.894	0.110
Var5	0.347	0.424	-0.707	0.894	**1**	0.144
Var6	-0.538	-0.597	-0.232	0.110	0.144	**1**

Table 4

Eigenvalues of the components

Principal Component Analysis:						
Eigenvalues:						
	F1	F2	F3	F4	F5	F6
Eigenvalue	3.045	2.134	0.429	0.296	0.087	0.009
Variability (%)	50.754	35.573	7.142	4.930	1.455	0.146
Cumulative %	50.754	86.327	93.468	98.398	99.854	100.000

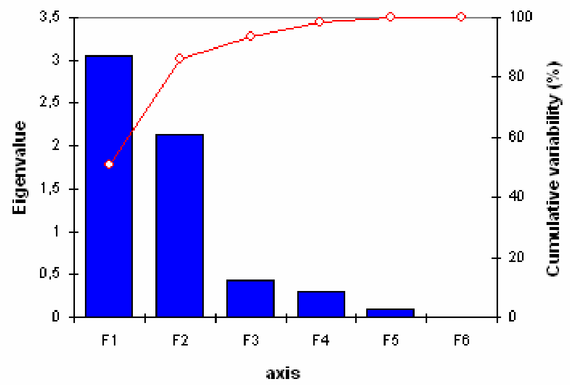

Fig. 1. Pareto type scree plot.

Table 5

Eigenvectors of the components

	F1	F2	F3	F4	F5	F6
Var1	0.389	-0.446	0.469	-0.221	0.197	-0.585
Var2	0.435	-0.422	0.205	0.059	-0.491	0.588
Var3	-0.411	-0.366	0.183	0.791	-0.089	-0.175
Var4	0.477	0.314	-0.324	0.307	-0.506	-0.468
Var5	0.498	0.254	0.103	0.478	0.622	0.249
Var6	-0.122	0.571	0.767	-0.003	-0.265	0.011

Table 6

Factor loadings of the components

	F1	F2	F3	F4	F5	F6
Var1	0.678	-0.652	0.307	-0.120	0.058	-0.055
Var2	0.760	-0.616	0.134	0.032	-0.145	0.055
Var3	-0.717	-0.535	0.120	0.430	-0.026	-0.016
Var4	0.832	0.458	-0.212	0.167	-0.149	-0.044
Var5	0.869	0.371	0.067	0.260	0.184	0.023
Var6	-0.213	0.834	0.502	-0.002	-0.078	0.001

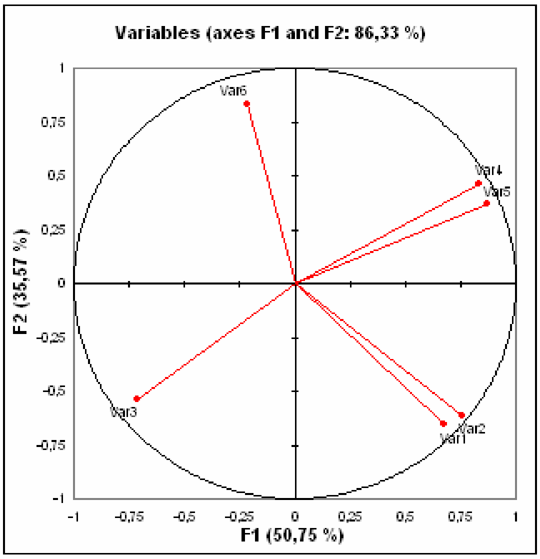

Fig. 2. Correlation circle.

The reduction is possible because the initial properties are related. The degree of correlations between the initial attributes and a few principal componenets is graphicaly described with the help of so-called correlation circle (Fig. 2). This map shows a projection of the initial variables in the factors space.

In the Table 8 the contribution of the variables (%) are given. Table 9 represents squared cosines of the variables.

There is no point in evaluating the number of PCs before. For the analysed data the first two PCs account for cca. 90% of the variance for the first five variables, meanwhile F_2 and F_3 explain approximately 90% of the variance for the sixth variable.

In the Pearson's method, each sample is considered as a point in the 6-dimensional space. Sample projections are called scores, variable projections are called loading. Table 10 contains coordinates of each object in the new 6-dimensional space, the complete principal components space.

Table 7

Correlations between variables and factors

	F1	F2	F3	F4	F5	F6
Var1	0.678	-0.652	0.307	-0.120	0.058	-0.055
Var2	0.760	-0.616	0.134	0.032	-0.145	0.055
Var3	-0.717	-0.535	0.120	0.430	-0.026	-0.016
Var4	0.832	0.458	-0.212	0.167	-0.149	-0.044
Var5	0.869	0.371	0.067	0.260	0.184	0.023
Var6	-0.213	0.834	0.502	-0.002	-0.078	0.001

Table 8

Contribution of the variables (%)

	F1	F2	F3	F4	F5	F6
Var1	15.116	19.889	22.029	4.868	3.865	34.232
Var2	18.962	17.777	4.212	0.346	24.103	34.602
Var3	16.869	13.403	3.34	62.523	0.79	3.074
Var4	22.747	9.847	10.494	9.442	25.576	21.895
Var5	24.819	6.459	1.059	22.82	38.659	6.183
Var6	1.487	32.626	58.867	0.001	7.006	0.013

Table 9

Squared cosines of the variables

Squared cosines of the variables:

	F1	F2	F3	F4	F5	F6
Var1	0.46	0.425	0.094	0.014	0.003	0.003
Var2	0.577	0.379	0.018	0.001	0.021	0.003
Var3	0.514	0.286	0.014	0.185	0.001	0
Var4	0.693	0.21	0.045	0.028	0.022	0.002
Var5	0.756	0.138	0.005	0.068	0.034	0.001
Var6	0.045	0.696	0.252	0	0.006	0

Table 10

Factor scores

Factor scores:

Obs.	F1	F2	F3	F4	F5	F6
Obs1	-0.814	3.615	-0.246	0.015	-0.224	-0.011
Obs2	-0.789	0.685	-0.559	0.729	0.433	0.052
Obs3	-1.06	-0.173	1.152	-0.261	-0.267	0.151
Obs4	-1.549	-1.319	0.624	0.928	-0.124	-0.124
Obs5	-0.805	-1.368	-0.453	-0.254	0.287	0.127
Obs6	-1.752	-1.194	-0.918	-0.667	-0.251	-0.09
Obs7	0.815	0.399	0.538	-0.769	0.293	-0.095
Obs8	2.483	0.026	0.39	-0.036	0.259	-0.052
Obs9	3.472	-0.67	-0.528	0.314	-0.405	0.044

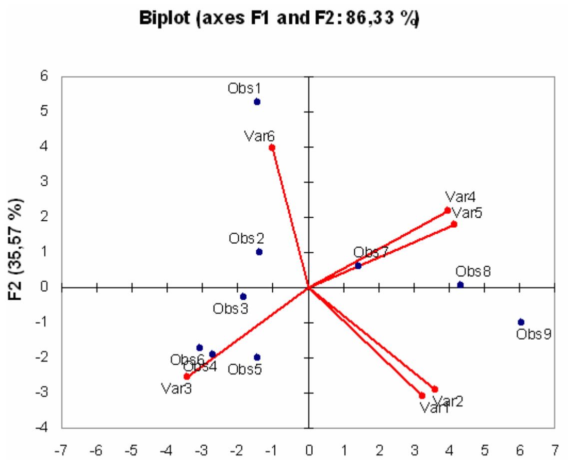

Fig. 3. Observations grouping on the factors axes.

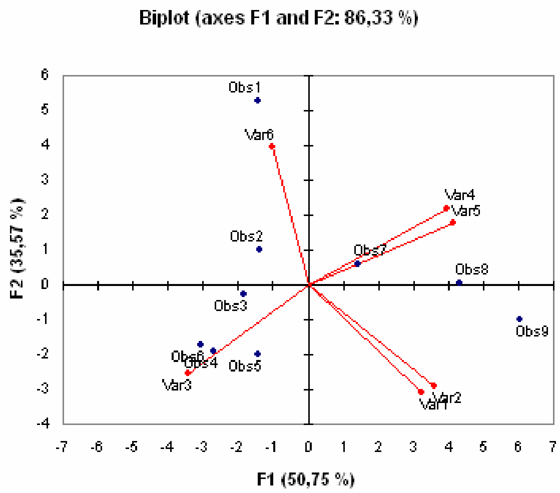

Fig. 4. Biplot graph.

Table 11

Contribution of the observations

Contribution of the observations (%):

	F1	F2	F3	F4	F5	F6
Obs1	2.416	68.015	1.565	0.009	6.364	0.151
Obs2	2.274	2.441	8.107	19.951	23.853	3.463
Obs3	4.101	0.155	34.429	2.565	9.096	28.74
Obs4	8.751	9.063	10.087	32.372	1.949	19.588
Obs5	2.365	9.741	5.324	2.426	10.445	20.34
Obs6	11.204	7.416	21.831	16.693	8.025	10.295
Obs7	2.424	0.828	7.496	22.222	10.898	11.526
Obs8	22.493	0.003	3.938	0.048	8.513	3.484
Obs9	43.972	2.337	7.223	3.714	20.857	2.412

Table 12

Squared cosines of the observations

Squared cosines of the observations:

	F1	F2	F3	F4	F5	F6
Obs1	0.048	0.944	0.004	0	0.004	0
Obs2	0.293	0.221	0.147	0.25	0.088	0.001
Obs3	0.425	0.011	0.502	0.026	0.027	0.009
Obs4	0.442	0.321	0.072	0.159	0.003	0.003
Obs5	0.224	0.648	0.071	0.022	0.028	0.006
Obs6	0.525	0.243	0.144	0.076	0.011	0.001
Obs7	0.369	0.088	0.161	0.329	0.048	0.005
Obs8	0.965	0	0.024	0	0.01	0
Obs9	0.924	0.034	0.021	0.008	0.013	0

The chart 3 allows visualizing and analyzes the observations, initially described by the 6 properties, on a 2-dimensional map, the optimal view for a variability criterion.

The biplot allows information on both samples and variables of a data matrix to be displayed graphically. Samples are displayed as points while variables are displayed either as vectors, linear axes or nonlinear trajectories.

In Table 11 are given the contribution of the observations (%).

Next the Table 12 represents squared cosines of the observations.

The explained validated variance is 86 % using 2 PCs. The two PCs of Joreskog technique explains about 95% of the variance, while the Pearson's procedure needed 3 PCs to explain the same variance.

3. CONCLUSIONS

Using PCA it is possible to describe a range of different materials in terms of principal attributes, which describe the relevant informations for the design. A possibility to reduce the number of attributes of materials in the design is the PCA model, based on Pearson' method.

In the present research it is used an artificial subspace with 2, respective 3 dimensions for the considerate range of materials. The real attributes of each object can be expressed with a good precision as function of artificial axes. The Pearson' method of PCA, which obtaints a few relevant factors from data sets, uses the XLSTAT 2009 software.

The application of this model will simplifies the materials design and there are many other possible extensions in the design process. Further more, a comparison between Jöreskog' and Pearson' method is presented for future developments.

REFERENCES

[1] Croux, C., Ruiz-Gazen, A. (2005). *High breakdown estimators for principal components: the projection-pursuit approach revisited*, Journal of Multivariate Analysis, Vol. 95, Issue 1 (July), pp. 206–226, ISSN:0047–259X.

[2] Jolliffe, I.T. (2002). *Principal Component Analysis*, Springer Verlag, 2nd ed., ISBN 978–0–387–95442–4.

[3] Hubert, M., Rousseeuw, P., Verboven, S. (2002). A *fast method for robust principal components with applications to chemometrics*, Chemometrics and Intelligent Laboratary Systems. Vol. 60.pp. 101–111.

[4] Maronna, R., A. (2005). *Principal Components and Orthogonal Regression based on Robust Scales,* Technometrics, 47, pp. 264–273.

[5] Pearson, K. (1901). *On Lines and Planes of Closest Fit to Systems of Points in Space*, Philosophical Magazine 2 (6), pp. 559–572.

[6] Shlens, J. (2005). *A Tutorial on Principal Component Analysis*, available at: `http://www. snl.salk.edu /~shlens/pub/notes/pca.pdf`, accessed: 2008-08-12.

[7] Târcolea, C., Paris, A. (2008). *The Joreskög technique applied for materials design*, Proceedings of the 17th International Conference on Manufacturing Systems – ICMaS, 13-14 nov. 2008, Edit. Academiei Române, ISSN 1842–3183, Romania, pp. 309–312.

[8] Târcolea, C., Paris, A., Târcolea-Demetrescu, A. (2009). *Statistical methods applied for materials selection.* The International Conference DGDS-2008 & MENP, 5 August 29 September, Mangalia, Romania, In Applied Sciences(APPS), ISSN 1454–5101, Vol. 11 (2009), Electronic Edition, pp. 145–150.

[9] Wilcox, R. (2008). *Robust principal components: A generalized variance perspective*, Behaviour Research Methods, 40 (1), pp. 102–108.

Authors:

PhD, Math., Constantin TÂRCOLEA, Professor, University "Politehnica" of Bucharest, Department of Mathematics,
E-mail: constantin_tarcolea@yahoo.com
PhD, Eng, Adrian PARIS, Lecturer, University "Politehnica" of Bucharest, Materials Tehnology and Welding Department,
E-mail: adrianparis@amza.camis.pub.ro