วิศวกรรมสาร มข. ปีที่ 32 ฉบับที่ 6 (793 - 801) พฤศจิกายน - ธันวาคม 2548
KKU Engineering Journal Vol 32 No 6 (793 - 801) November - December 2005

วิธีวิเคราะห์การจัดลำดับรางวัลโดยใช้ส่วนประกอบสำคัญของปัจจัย*

ชาญณรงค์ สายแก้ว[1)]

[1)] ผู้ช่วยศาสตราจารย์ ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น จังหวัดขอนแก่น 40002

Email: charn_sa@kku.ac.th

บทคัดย่อ

จุดประสงค์ของการศึกษานี้เพื่อ เสนอวิธีการจัดลำดับรางวัลในการประกวดการแข่งขันผลงานงานโครงการของนักศึกษาชั้นปีสี่ซึ่งจัดโดยคณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่นเป็นประจำทุกปี งานโครงการได้รับการคัดเลือกจากภาควิชาต่างๆ เพื่อเป็นตัวแทนในการประกวดรับรางวัลประเภทต่างๆ อาทิ ประเภทสิ่งประดิษฐ์ ประเภทการประยุกต์ใช้งาน และประเภทความรู้พื้นฐานทางด้านวิศวกรรม วิธีวิเคราะห์โดยใช้ส่วนประกอบสำคัญของปัจจัยซึ่งเป็นวิธีการวิเคราะห์ทางสถิติหลายตัวแปรถูกนำมาใช้ในการวิเคราะห์ข้อมูลที่เป็นค่าเฉลี่ยของคะแนนที่ได้รับจากคณะกรรมการตัดสินในแต่ละเกณฑ์การให้คะแนนเพื่อจัดลำดับการให้รางวัล

คำสำคัญ : การจัดลำดับรางวัล, วิธีวิเคราะห์โดยใช้ส่วนประกอบสำคัญของปัจจัย

---

* รับต้นฉบับเมื่อวันที่ 5 พฤษภาคม 2548 และได้รับบทความฉบับแก้ไขเมื่อวันที่ 9 กรกฎาคม 2548

An Approach for Prize Ranking Analysis Using Principal Components*

Charnnarong Saikaew, Ph.D. [1)]

[1)] Lecturer, Department of Industrial Engineering, Faculty of Engineering
Khon Kaen University, Khon Kaen 40002
Phone: +66-43-343117, Fax: +66-43-343117

Email: charn_sa@kku.ac.th

Abstract

The objective of this study is to provide a methodology for ranking the selected senior engineering projects for earning prizes awarded annually by Faculty of Engineering at Khon Kaen University. The senior engineering projects are selected from all departments as candidates for outstanding awards in the innovation category, engineering application category, and basic engineering knowledge category. For defining a prize ranking, the method of principal component analysis known as a multivariate statistical technique is used to perform analysis on the average data sets of earning scores judged on the referees' evaluations for each of the appropriate criteria.

Keywords : Prize Ranking, Principal Component Analysis

---

* Original manucript submitted: May 5, 2005 and Final manucript received: July 9, 2005

Introduction

During academic years, senior engineering projects are selected as candidates for outstanding awards in the innovation category, engineering application category, and basic engineering knowledge category. The prizes are awarded annually to the winners who are the first, second, and third in the senior project competition event held by the Faculty of Engineering at Khon Kaen University. The existing methodology used to rank the projects was based on a total score of each project as well as the faculty referees' judgments. However, the referees judged on their opinions to the quality of the projects, which were too subjective in that they lacked a valid quantifiable methodology. Hence this study proposes a methodology for prize ranking analysis based on an objective assessment of the average data sets of earning scores judged on the referees' evaluations for each of the appropriate criteria. To meet this purpose, the method of principal component analysis (PCA) is used to perform analysis on such the data sets.

The paper organization begins with principal component analysis, which provides a literature review on the background and theory of PCA procedure. The last two sections present the results of the prize ranking scheme used for awarding the prize winners and give a summary of the results, respectively.

Principal Component Analysis

Principal component analysis (PCA) is a multivariate approach in which a number of related variables are linearly transformed to a set of uncorrelated variables (Jackson, 1991). The procedure of PCA is to transform a vector of the original correlated variables, $[x_{11}\ x_{12} \ldots x_{1p}]$, into a vector of the new variables, $[z_{11}\ z_{12} \ldots z_{1p}]$, that are uncorrelated with each other. The vector of the new variables represents the selection of a new coordinate system obtained by rotating the original coordinate system with $x_{11}, x_{12}, \ldots, x_{1p}$ as the coordinate axes. The logic of rotating the original coordinate system is to maximize the variance of the new coordinate system. After the first coordinate axis on which the variance is maximal, there remains some variability around this coordinate axis. In principal component analysis, after the first coordinate axis has been extracted (i.e., after the first coordinate axis has been drawn through the data), another coordinate axis will be extracted in order to maximize the remaining variability, and so forth. In this manner, consecutive coordinate axes are extracted. Since each consecutive coordinate axis is defined to maximize the variability that is not captured by the preceding coordinate axis, consecutive coordinate axes are independent of each other. This implies that consecutive coordinate axes are uncorrelated or orthogonal to each other. The original data can be transformed into the new coordinate axes by this manner.

Let $\mathbf{x}_k$ be a vector of the original independent variables for each observation k defined as $[x_{k1}\ \ x_{k2} \ldots x_{kp}]$, $k = 1, 2, \ldots, n$. Let $\mathbf{z}_k$ be a vector of principal component scores of $\mathbf{x}_k$ for observation k defined as $[z_{k1}\ \ z_{k2} \ldots z_{kp}]$, $k = 1, 2, \ldots, n$ where z_{kp} is a principal component score of $\mathbf{z}_k$. Let $\mathbf{U} = [\mathbf{u}_1\ \ \mathbf{u}_2 \ldots \mathbf{u}_p]$. Thus, the orthogonal transformation for each observation, k, is defined as follows:

$$\mathbf{z}_k = \mathbf{x}_k \mathbf{U}, \quad k = 1, 2, \ldots, n \tag{1}$$

where $\mathbf{U}$ is an orthogonal matrix $p \times p$ (i.e., $\mathbf{U}'\mathbf{U} = \mathbf{I}$ and $\mathbf{U}\mathbf{U}' = \mathbf{I}$, where $\mathbf{I}$ is identity matrix). The coordinate axes of the uncorrelated variables are described by the vectors $\mathbf{u}_i$ that make up the matrix $\mathbf{U}$ of direction cosines. The columns of $\mathbf{U}$ are called eigenvectors. A first column vector of $\mathbf{U}$ defined as $\mathbf{u}_1$ or $[u_{11}\ u_{21} \ldots u_{p1}]'$ represents the coefficients of the first principal component (pc) called the first eigenvector of the covariance or correlation matrix, a second vector $\mathbf{u}_2$ or $[u_{12}\ u_{22} \ldots u_{p2}]'$ represents the coefficients of the second pc called the second eigenvector, and the last vector $\mathbf{u}_p$ or $[u_{1p}\ u_{2p} \ldots u_{pp}]'$ represents the coefficients of the p pc called the p eigenvector. Thus, principal components are particular linear combinations of the p original random variables with coefficients equal to the eigenvectors of the covariance or correlation matrix.

A $p \times p$ covariance matrix **S** can be reduced into a diagonal matrix **L** by premultiplying and postmultiplying it by a particular orthogonal matrix **U**. Therefore,

$$\mathbf{U'SU} = \mathbf{L} \tag{2}$$

where

$$\mathbf{L} = \begin{bmatrix} l_1 & 0 & \cdots & 0 \\ 0 & l_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & l_p \end{bmatrix}. \tag{3}$$

Here **S** is the covariance matrix of the original variables while matrix **L** is the covariance matrix of the principal components. The diagonal elements of matrix **L** are called eigenvalues, which represent the variances of the new variables. The off-diagonals of the matrix **L** representing the covariance are zero. This implies that $z_{11}, z_{12}, \ldots, z_{1p}$ are uncorrelated. The trace of the covariance matrix **L** is equal to the sum of the original variances; this is known as one of the important properties of PCA. According to this property, the percentage or proportion of the total variability is accounted for by each *pc* as defined as follows:

$$pr_i = \frac{l_i}{\sum_{i=1}^{p} l_i} \times 100 \tag{4}$$

where pr_i is the percentage of the total variability accounted for by each *pc*. This value is important for *pc* interpretations. It is worth noting that the principal components are sorted by descending order of eigenvalues (i.e., from the largest variance of any linear combination of the original variables to the smallest variance of any linear combination of the original variables). This implies that the first principal component accounts for as much variation in the data as possible while each succeeding principal component accounts for as much of the variation unaccounted for by preceding principal components as possible.

Principal component analysis can be applied to a wide variety of problem ranking analyses. An application of PCA can be found in the study of the Olympic track record ranking by Dawkins (1989). Dawkins used PCA to perform a national track record ranking study of the fifty-five countries for men's and women's events. Since the time units of the data sets were different, each data set needed to be scaled (i.e., normalized with a mean zero and unit standard deviation). Since the first principal component accounted for maximum variation, it was reasonable to use the first principal component for ranking the countries. In 1996, Naik and Khattree revisited the Olympic track record ranking where PCA was utilized. In order to compare the athletic performances of the nations, Naik and Khattree (1996) stated that the appropriate variables were the speeds rather than the total time taken because these variables succeeded in retaining the possibility of having different degrees of variability in different variables. The speed variables were defined as the distances (in meters) covered per second for the various track events, which were in the same unit. Thus, unlike the starting point of PCA with the correlation matrix used in Dawkins's study, Naik and Khattree (1996) used the covariance matrix to perform the national track record ranking study. The PCA results showed that the first principal component represented a weighted average of all the speeds in the various events, which measured the overall athletic excellence among the nations whereas the second principal component appeared to be interpretable as the measure of differential achievement. Another application of PCA can be found in the study of the corrosion severity ranking by location of the six United States operational air bases by Saikaew and Court (2004). The six air bases worldwide were chosen by the United States Air Force (USAF) to represent the range of atmospheric conditions that C/KC-135 aircraft could be exposed to over their operational life. A corrosion severity ranking scheme for the six air bases allowed the USAF to concentrate their efforts on proactively inspecting aircraft for corrosion when deployed and operated at highly severe corrosion sites. The method of PCA was used to analyze compositional data sets of atmospheric conditions for defining corrosion severity ranking by locations (air bases). The first two principal components were used for corrosion severity ranking analysis since the first two principal components could account for most of the variability of the data. Saikaew (2005) proposed a methodology for industrial problem ranking analysis in cleaner technology practices based on an objective assessment of the input resources, raw materials, production capacity, as well as waste generated by the industrial process. The severity industrial problem(s) would allow the practitioners to concentrate their efforts on implementing the CT concept for solving such the problem(s). The method of principal component analysis was used for simultaneously analyzing the compositional data sets of the input resources, raw materials, production capacity, and waste for defining the industrial problem ranking.

Even though PCA can be applied to a wide variety of problem ranking analyses as discussed thus far, it aims to be flexible approach which can be equally applied to wider contexts. Hence PCA can be applied to the analysis of prize ranking scheme.

Results and Discussion

During the academic year 2004-2005, seventeen senior projects were selected as candidates for outstanding awards in the innovation category. Data sets of the seventeen projects are collected based on scores earned with the nine criteria used for judging prize ranking. The nine criteria selected as independent variables for defining prize ranking include 1) *criterion*1: creative ideas, 2) *criterion*2: scientific skill, 3) *criterion*3: applications, 4) *criterion*4: oral presentation skill, 5) *criterion*5: knowledge and understanding in context, 6) *criterion*6: project contribution, 7) *criterion*7: context type, 8) *criterion*8: designing a poster presentation, and 9) *criterion* 9: context of the poster. Full scores for *ctiterion*1: 25, *criteria*2, 4, 5, 6, and 7: 10, *criterion*3: 15, and *criteria*8 and 9: 5. Table 1 shows the data sets of the seventeen projects with the nine criteria, an average score for each criterion, and a total score for each project. The results of the principal component analysis illustrate the covariance matrix, eigenvalues of covariance matrix, a scree plot of eigenvalues, eigenvectors, the first two principal components, and a scatter plot of the first two principal components for the scores earned with the nine criteria. These results are illustrated in Tables 2-5 and Figures 1 and 2.

	*Crit*1	*Crit*2	*Crit*3	*Crit*4	*Crit*5	*Crit*6	*Crit*7	*Crit*8	*Crit*9	Total
Project01	18.67	7.83	11.67	6.33	7.00	7.50	7.00	3.17	3.08	72.25
Project02	18.60	8.00	10.60	7.00	8.00	7.67	7.83	3.25	3.08	74.03
Project03	15.60	7.40	10.80	6.60	6.60	6.50	6.50	2.50	2.67	65.17
Project04	20.17	8.50	11.83	8.83	8.83	7.83	8.17	3.00	3.50	80.66
Project05	19.00	8.17	12.17	8.17	7.83	7.67	7.67	3.50	3.75	77.93
Project06	17.71	7.43	11.29	6.71	6.71	6.43	6.79	2.29	2.29	67.65
Project07	21.71	8.83	13.43	8.71	8.86	8.57	8.43	3.00	2.86	84.40
Project08	21.43	8.43	12.57	8.29	8.29	8.71	8.57	3.21	3.00	82.50
Project09	20.38	8.13	12.63	9.00	8.75	8.88	8.50	3.19	3.31	82.77
Project10	18.75	7.75	12.50	8.75	8.50	8.00	7.50	4.10	4.30	80.15
Project11	19.25	8.00	12.75	8.33	8.75	8.00	8.25	3.80	4.10	81.23
Project12	17.50	7.75	11.75	8.25	8.50	7.75	7.25	4.40	4.10	77.25
Project13	15.33	6.33	10.33	7.67	7.67	6.75	6.25	3.75	3.75	67.83
Project14	21.67	8.67	12.00	7.67	8.00	8.75	8.50	3.00	3.25	81.51
Project15	22.20	9.00	13.20	8.60	8.40	9.00	9.00	3.58	3.58	86.56
Project16	20.40	8.60	12.00	8.60	8.60	8.20	8.40	3.25	3.58	81.63
Project17	21.75	7.50	12.25	7.75	7.50	7.80	7.80	3.00	2.50	77.85
Average	19.42	8.02	11.99	7.96	8.05	7.88	7.79	3.29	3.34	

Table 1 *Data sets*

	*Crit*1	*Crit*2	*Crit*3	*Crit*4	*Crit*5	*Crit*6	*Crit*7	*Crit*8	*Crit*9
*Crit*1	4.3605								
*Crit*2	1.1110	0.4281							
*Crit*3	1.4241	0.4027	0.7500						
*Crit*4	0.9117	0.2703	0.5002	0.7138					
*Crit*5	0.7958	0.2727	0.3892	0.5634	0.5395				
*Crit*6	1.4022	0.4088	0.5351	0.4605	0.4310	0.6155			
*Crit*7	1.5176	0.4645	0.5341	0.4514	0.4259	0.5828	0.6418		
*Crit*8	-0.0986	-0.0228	0.0833	0.2077	0.2073	0.1109	0.0292	0.2801	
*Crit*9	-0.1416	0.0172	0.0871	0.2668	0.2525	0.1151	0.0575	0.2723	0.3312

Table 2 *Covariance matrix*

Number	Eigenvalue	Difference	Proportion	Cumulative
1	6.692137	5.434099	0.7727	0.7727
2	1.258038	1.035082	0.1453	0.9179
3	0.222956	0.030185	0.0257	0.9437
4	0.192771	0.030938	0.0223	0.9659
5	0.161833	0.101447	0.0187	0.9846
6	0.060386	0.023612	0.0070	0.9916
7	0.036774	0.013782	0.0042	0.9958
8	0.022992	0.009892	0.0027	0.9985
9	0.013100		0.0015	1.0000

Table 3 *Eigenvalues of the covariance matrix*

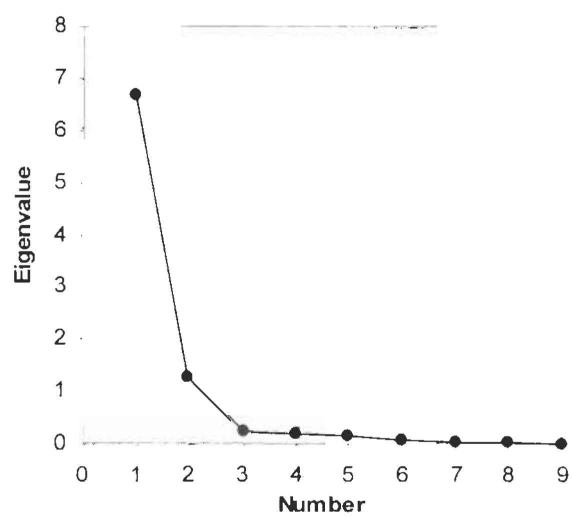

Figure 1 *Scree plot*

As mentioned before, eigenvalues are ordered from largest to smallest with the first few explaining most of the variability. Generally, the scree plot involves selecting the number of important principal components based on the visual appearance of the plot. The scree plot illustrates a steep drop over the first few components followed by a leveling off for the rest of the components. The criterion of this test is to plot a graph of eigenvalues as the function of the number of variables and try to draw a straight line connecting all eigenvalues. The number of retained variables can be found if all components have eigenvalues above the straight line. According to Figure 1, the number (2) of principal components to be selected is determined in such a way that the slope of the plot is steep to the left of 2 but at the same time not steep to the right.

According to Tables 3 and 4, the first principal component with the variance 6.6921 explains about 77 percent of the variability and the eigenvector corresponding to the variance 6.6921 listed under the column $\mathbf{u}_1$ is $(0.7881, 0.2161, 0.2864, 0.2167, 0.1905, 0.2804, 0.2963, 0.0107, 0.0118)'$. That is, the first principal component with all its positive coefficients seems to measure the general ranking measurement of a project. The second principal component with the variance 1.2580 explains about 15 percent of the variability and the eigenvector corresponding to the variance 1.2580 listed under the column $\mathbf{u}_2$ is $(-0.3694, -0.0229, 0.1306, 0.4975, 0.4421, 0.1441, 0.0552, 0.3973, 0.4701)'$. That is, the second principal component consists of positive and negative coefficients indicating a contrast between the first group of the considered variables (criteria 1 and 2) and the second group of the considered variables (criteria 3-9). By examining the cumulative proportion of the variation given in Table 3 and scree plot depicted in Figure 1, at least two principal components are needed to account for 92 percent (77 + 15) of the total variability. This implies that the first two principal components can account for most of the variability of the data.

	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$	$\mathbf{u}_4$	$\mathbf{u}_5$	$\mathbf{u}_6$	$\mathbf{u}_7$	$\mathbf{u}_8$	$\mathbf{u}_9$
Crit1	0.7881	-0.3694	0.0204	-0.1575	-0.3738	-0.2594	-0.0176	-0.0556	-0.0810
Crit2	0.2161	-0.0229	0.2615	0.0897	0.7276	-0.3510	0.0420	-0.4430	0.1607
Crit3	0.2864	0.1306	-0.8218	-0.0958	0.3903	0.1133	-0.1650	0.1437	-0.0564
Crit4	0.2167	0.4975	-0.1866	0.5530	0.2717	-0.0678	0.4497	-0.1884	0.2203
Crit5	0.1905	0.4421	0.2419	0.2878	-0.0605	0.0209	-0.7383	-0.0154	-0.2780
Crit6	0.2804	0.1441	0.2181	-0.2288	0.1027	0.7708	0.2461	-0.2890	-0.2274
Crit7	0.2963	0.0552	0.3190	0.1101	0.2304	0.1615	0.0825	0.7494	0.3841
Crit8	0.0107	0.3973	0.0107	-0.5943	0.1794	-0.0536	-0.2029	-0.1779	0.6170
Crit9	0.0118	0.4701	0.1136	-0.3892	0.0708	-0.4103	0.3387	0.2595	-0.5087

Table 4 *Eigenvectors*

Project	Principal component 1	Principal component 2
Project01	-1.6191	-1.3002
Project02	-1.3138	-0.5404
Project03	-4.8392	-0.9431
Project04	1.0867	0.4462
Project05	-0.3276	0.4255
Project06	-2.9250	-1.8121
Project07	3.0869	-0.1475
Project08	2.4184	-0.4310
Project09	1.8151	0.6867
Project10	-0.2123	1.6907
Project11	0.4805	1.2626
Project12	-1.6642	1.7810
Project13	-4.9599	1.1520
Project14	2.2977	-1.0004
Project15	3.6367	0.0418
Project16	1.4202	0.3681
Project17	1.6188	-1.6800

Table 5 *The first two principal components*

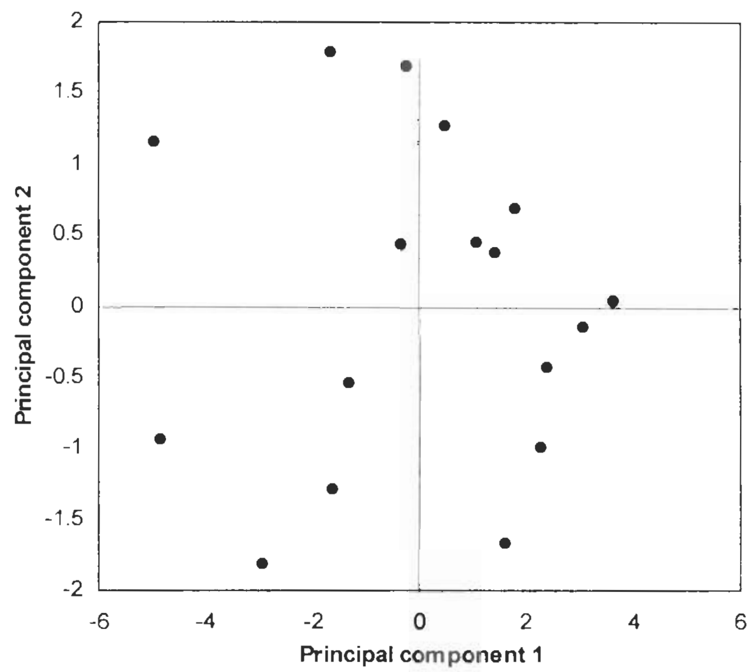

Figure 2 *Scatter plot of the first two principal components for the seventeen projects*

The first two principal component equations for each of the seventeen projects are given by:

$$pc_1 = 0.7881\,Crit1 + 0.2161\,Crit2 + 0.2864\,Crit3 + 0.2167\,Crit4 + 0.1905\,Crit5 + 0.2804\,Crit6 + 0.2963\,Crit7 + 0.0107\,Crit8 + 0.0118\,Crit9$$

and

$$pc_2 = -0.3694\,Crit1 - 0.0229\,Crit2 + 0.1306\,Crit3 + 0.4975\,Crit4 + 0.4421\,Crit5 + 0.1441\,Crit6 + 0.0552\,Crit7 + 0.3973\,Crit8 + 0.4701\,Crit9$$

where pc_1 and pc_2 are the first and the second principal components.

Since eigenvectors are orthogonal, the principal components represent jointly perpendicular directions through the space of the original variables. This implies that the transformed variables are uncorrelated with each other and centered at the origin. Based on the scatter plots of the first two principal components depicted in Figure 2, the horizontal line and the vertical line aligned zeros are used to separate the projects into four quadrants. The northeast quadrant is considered as the most favorable zone for the appropriate prize-winning projects while the southwest quadrant is considered as the least favorable zone. Therefore, the results of the most severe and least favorable zones as determined by the scatter plot of the two principal components can be used to rank the prize-winning projects. Since the first principal component accounts for variability more than the second principal component, the remaining projects lying in the northwest and the southeast should rank by considering the first principal component. Accordingly, the ranking for the seventeen projects ranks from the most favorable project to the least favorable project: Projects15, 09, 16, 04, 11, 07, 08, 14, 17, 10, 05, 12, 13, 02, 01, 06, and 03.

Based on the three scenario considerations, Table 6 shows the results of prize ranking of the prize-winning projects. All three scenarios show that the most favorable project is Project15 whereas the remaining rankings are different. Typically, an advantage in use of PCA for ranking analysis is that the first few principal components can account for most of the variability of the data while the other two ranking schemes are not concerned with explaining variability.

Rank	Official judging	Total score	PCA
1st Place	Project15	Project15	Project15
2nd Place	Project09	Project07	Project09
3rd Place	Project08	Project09	Project16

Table 6 *Prize ranking of the three scenarios based on official judging, total score, and PCA*

Conclusion

Principal component analysis is a flexible approach, which can be applied to a wide variety of problem ranking analyses including the analysis of prize ranking scheme. Since the first few principal components account for a large proportion of the variability, this achieves the objective of ranking analysis which is a better evaluation than simple averages of the original variables and the referees' judgments (subjective evaluation). Thus, principal component analysis can be useful when there is a high degree of correlation present in the multivariate.

The applications of principal component analysis depend on data availability. The extractions of principal components can be made using either original multivariate data sets or using the covariance or the correlation matrix. Generally, the correlation matrix is used when the original variables are measured in different units or have much different variances. More details on the data scaling in PCA can be found in Jackson (1991) and the references therein.

References

Dawkins, B. 1989. Multivariate Analysis of National Track Records. **The American Statistician**, 43(2), 110-115.

Jackson, J.E. 1991. **A User's Guide to Principal Components**. John Wiley & Sons, Inc., New York.

Naik, D.N. and Khattree, R. 1996. Revisiting Olympic Track Records: Some Practical Considerations in the Principal Component Analysis. **The American Statistician**, 50(2), 140-144.

Saikaew, C. and Court, M.C. 2004. A Corrosion Severity Ranking Methodology Based on Environmental Data. **Proceedings of the 33rd International Conference on Computers and Industrial Engineering**, Jeju, Korea, March 25-27.

Saikaew, C. 2005. A Proposed Approach for Evaluating the Industrial Problem Ranking in Cleaner Technology Practices. **Proceedings of the 35th International Conference on Computers and Industrial Engineering**, Istanbul, Turkey, June 19-22.