Determination of the Rankings of the Teams Competed in the 2009 – 2010 BEKO Basketball League with Principal Components Analysis

Gürol ZIRHLIOĞLU, Murat KAYRİ and Mustafa ATLI [+]

Yuzuncu Yil University, Education Faculty, Van / TURKEY

Abstract. Being parallel with the technological advancements, statistical analysis methods have a wide area of usage in many fields of science. Being started to be used particularly in the recent years also in the sciences of sports, statistical analysis methods have been proved to be an important factor in obtaining more positive results in evaluating either team or individual performances. There are many factors that affect performance and success in all areas of sports. Sometimes examining all these factors together may prove to be a problem. Factor analysis is one of the multivariate statistical methods that transform great numbers of interrelated variables into a few factors.

In this present paper, an alternative ranking among the teams that competed in the BEKO basketball league 2009-2010 season has been derived by means of utilizing the teams' match performances with the factor analysis approach. At the end of the study number of the factors determining the ranking has been reduced to two and an alternative ranking based on the 1st factor with the biggest variance explanation rate has been established. According to this new ranking, the team Fenerbahce Ulker, which was in the 2nd rank in the original ranking, has taken the first place.

Keywords. Principal components analysis, dimension reduction, BEKO basketball league

1. Introduction

With the advancing technology, statistical analyses are finding ever wider areas of usages in many different fields of science such as Psychology, Sociology, Medicine, Agriculture, Biology, etc. Statistical analysis methods have been started to be widely used also in sports sciences. The wider usage of statistical methods in most sports branches, with football, basketball and volleyball taking the first places, appears as a result of these advancements.

Evaluation of teams and individual sportspersons on the basis of their performances may prove to be more important than the evaluation made by only considering match results. Performance evaluations made either during or after the match such as the mistakes, inadequacies, correct or wrong moves of the teams or individuals, selection of the right player at the right time, etc. can be achieved with a set of statistical indicators of the teams or players (Zirhlioglu and Karaca, 2005; Cengiz and Kilinc, 2007).

Considering statistical approaches for both team-based and individual performance evaluations also brings along a set of positive results. The statistical methods used in sports sciences are generally in one of the forms of descriptive statistical methods, simple correlation calculations or cross-tables. This however can cause many features of the data to be overlooked.

There are many factors in a sports match that may affect success or failure. These factors may originate from the players, teams, coaches, training programs, performance displayed throughout the match and environmental factors. There are many factors that affect performance and success in all areas of sports (İnce and Şen, 2006).

Factor analysis is one of the commonly used multivariate statistical methods that transform several interrelated variables into a few, easily understandable and more significant, independent factors. Factor analysis aims to obtain the unobserved major and minor factors that reflect the classification, and that are

---

[+] Corresponding author. Tel.: +90-554-300 6570.
 E-mail address: gurol@yyu.edu.tr

generated with the gathering of the variables, from the observed and correlated p variables in the x data matrix (Özdamar, 2010).

In factor analysis, there is no dependent variable and independent variables trying to explain this variable, as it is the case in regression analysis. In factor analysis, constitution of general variables called factors by bringing together variables that have high levels of correlation between them is in question. Main objectives of factor analysis are to reduce the number of variables that explain the dependent variable and to set forth the structure in the relations between the variables, in other words to classify the variables (Johnson and Wichern, 2007; Cengiz and Kılınç, 2007; Özdamar, 2010).

Factor analysis can be referred to with several terms such as explanatory factor analysis, confirmatory factor analysis, Q type, R type, O type, T type and S type factor analysis methods, depending on its way and purpose of implementation.   With a general classification factor analyses can be categorized under two groups as explanatory (exploratory) and confirmatory factor analyses. Explanatory factor analysis is the type of factor analysis that is used for deriving a fewer number (k<p) of independent new variables from a data set consisting of p number of interrelated variables by utilizing variance-covariance or correlation matrix. While variance-covariance is used for the variables included in the model as raw data, for variables transformed into a standard (such as z, t, log) correlation matrix is utilized.

Beyond reducing variables and renaming the newly generated factors, explanatory factor analysis examines whether the indicators gathered under a given factor are the indicators of the theoretical structure (Çokluk et al., 2010; Ozdamar, 2010). On the other hand, confirmatory factor analysis is the factor analysis used to test the conformity of the factors determined with explanatory factor analysis to presumptive or theoretical factor structures (Ozdamar, 2010). Explanatory factor analysis is mostly used for finding and setting forth the implicit sources of the variance and covariance in the observed measurements. For this reason explanatory factor analyses can play a considerably important role in developing tests (Joreskog and Sorbom, 1993).

Since the internal relations between the variables are determined in factor analysis the availability of a certain level of multicollinearity is desired. Having an adequate level of significance of the correlation between the variables and having correlation coefficients larger than 0.30 are the expected conditions in order to implement factor analysis (Hair et al., 1998).

In this present paper, it has been tried to set forth that an alternative ranking among the teams that competed in the BEKO basketball league 2009-2010 season can be derived by means of utilizing the teams' match performances with factor analysis approach.

2. Method

In the study carried out the data of 16 teams that competed in the 2009-2010 season of BEKO Basketball league have been utilized. These data have been collected from the official web site of Turkish Basketball Federation. The values of the variables *total points* (SY), *field goal attempts (2SY), field goals made* (2SYIS), *three points shot attempts (3SY), three points shots made* (3SYIS), *free throw attempts (1SY), free throws made* (1SYIS), *offensive rebounds* (HR), *defensive rebounds* (SR), *total rebounds* (TR), *assists* (AS), *blocks* (BL), *steals* (TC), *turnovers* (TK) and *fouls* (FA) indicated for each team in the web site have been used. The scored points based rankings of the teams competed in the 2009-2010 BEKO basketball league have been reevaluated by means of factor analysis.

Kaiser-Mayer-Olkin (KMO) conformity test value has been considered in order to asses the conformity of the data set to factor analysis. KMO test is a value that compares observed correlation coefficients with partial correlation coefficients. Value of the test varies between 0-1. KMO value is obtained with the following equation

$$KMO = \frac{\sum\sum\limits_{i \neq j} r_{ij}^2}{\sum\sum\limits_{i \neq j} r_{ij}^2 + \sum\sum\limits_{i \neq j} a_{ij}^2}$$

In the formula, while r_{ij} shows the simple correlation coefficient between the variables i and j, a_{ij} shows the partial correlation coefficient between the variables i and j. It is decided that the data are not conforming to factor analysis in case where the obtained KMO value is below the value 0.50. For an outstanding factor analysis KMO value should be greater than 0.80, still values greater than 0.50 can also be accepted.

2.1. Factor Model

Between the factors that can not be observed with the X observation vector, two types of factor models can be established as oblique and orthogonal.

Oblique factor model aims to determine the factors by presuming that there is k number of unobserved common factors dependent to X in a curvilinear way (F_1, F_2, …, F_k) and that there are p number of special factors. In other words, oblique factor model is used when the relation between the variables is nonlinear.

On the other hand, orthogonal factor model aims to determine the factors by presuming that there are k number of unobserved common factors that are linearly dependent with X but independent between each other (F_1, F_2,.., F_k) and that there are p number of special factors named as error. In here, the relation between the independent variable and predicted variables is seen as linear. Factor analysis model can be written as follows:

$$X_1 - \mu_1 = l_{11}F_1 - l_{12}F_2 + ... + l_{1k}F_k + e_1$$
$$X_2 - \mu_2 = l_{21}F_1 - l_{22}F_2 + ... + l_{2k}F_k + e_1$$
$$\vdots \qquad \qquad \vdots$$
$$X_p - \mu_p = l_{p1}F_1 - l_{12}F_2 + ... + l_{pk}F_k + e_p$$

In here, the 1_{pk} coefficient determines the load of the p. variable on k. factor. Rewriting the above factor analysis model in matrix form can be as follows:

$$X - \mu = L\,F + e$$

In here, X shows the -μ (px1) dimensioned difference vector, L shows the (pxk) dimensioned factor loads matrix, F shows the (kx1) dimensioned factor vector and e shows the (px1) dimensioned error vector (Johnson and Wichern, 2007).

Since the study conducted had the purpose of determining the unrelated linear components between the variables of the observations and of making a new ranking according to these components, the principal components (basic components) method has been used in determining the factors.

Principal components model aims the explained variance of the variable to be maximum. In the model, the principal independent components of equal in number to the number of variables, make maximum contribution to the total variance where all variables are explained. In other words, it is the method for expressing the variance structure of the original p variables with fewer variables that are the linear components of this variables. At the end p number of variables are represented by again p number of independent principal components and these components have maximum contribution to the total variance. In a sense, principal components join the variance at maximum, while other components do so at ever decreasing amounts (Rencher, 2002; Özdamar, 2010).

2.2. Determination of number of factors

In order to determine the necessary number of factors, it has to be examined what percent of the total variance is explained by each factor. Kaiser and Cattel scree test criteria are frequently used in order to determine the number of factors according to the variance the factors explain. It is a commonly used criteria to determine factors as much as number of radixes greater than one ($\lambda \geq 1$) in the S or R matrix. In the Cattel Scree test, a line curve graphic is drawn with number of components being at the x axis and eigenvalues being at the y axis. In the graphic, the point where the slope starts to disappear is expressed as the factor number. After this point, the contributions made by the components will be both smaller and approximately the same (Özdamar, 2010; Çokluk et. al., 2010).

3. Results

3.1. Kaiser-Mayer-Olkin (KMO) Test

The value obtained from the KMO test, conducted in order to determine whether the variables are conforming to factor analysis, has been found out to be 0.720. This indicates a high level of correlation among the variables. Therefore it has been decided that the data set was suitable for factor analysis.

3.2. The number of factors

The number of factors to be derived has been determined by using participation to variance (eigenvalue) criteria and Scree test criteria (mass graphic). Results obtained as per participation to variance criteria are as

shown in Table 1.

Table 1. Number of factors according to variance participation criteria

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Cumulative %	% of variance	Cumulative %
1	10.673	71.154	71.154	10.673	71.154	71.154	7.715	51.433	51.433
2	1.176	7.842	78.996	1.176	7.842	78.996	4.134	27.563	78.996
3	0.905	6.036	85.032						
4	0.607	4.044	89.076						
5	0.515	3.434	92.510						
6	0.396	2.641	95.151						
7	0.329	2.195	97.346						
8	0.229	1.528	98.874						
9	0.117	0.778	99.652						
10	0.026	0.172	99.824						
11	0.018	0.118	99.943						
12	0.005	0.035	99.978						
13	0.003	0.022	100.000						
14	4.84E-005	0.000	100.000						
15	6.47E-007	4.31E-006	100.000						

The components of which eigenvalue statistic is larger than 1, according to the variance participation criteria shown in Table 1, express the total factor numbers. In this study there have been 2 factors with eigenvalue greater than 1, and these 2 factors explain 78.996 % of the total variance.

The mass graphic on the number of factors to be derived as per the scree test criteria has been obtained as shown in Figure 1.
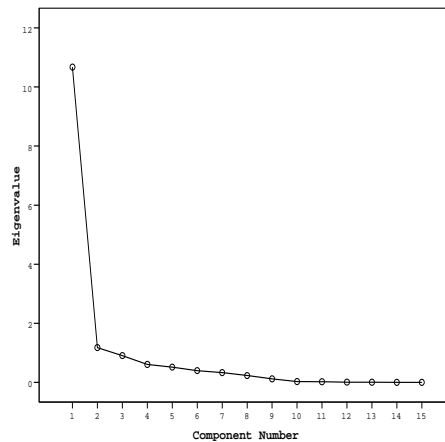

Figure 1. Factor analysis mass graphic

Generating factors has been ceased at the point where the slope becomes horizontal in the mass graphic given in Figure 1. The point where the slope starts to become horizontal indicates the point where variance participation value starts to be less than 1. In this case it is seen that the number of factors to be derived is 2.

3.3. Generation of factors

Matrix of the rotated factor values of factor generation has been obtained as shown in Table 2.

As indicated in Table 2, while variables of defensive rebounds, total points, total rebounds, field goal attempts, field goals made, three points shot attempts, three points shots made, free throw attempts, free throws made and assists are taken in the first factor, variables of steals, offensive rebounds, turnovers, fouls and blocks are taken in the second factor.

Table 2. Rotated factor values

Variables	Factors	
	1	2
SR	**0.928**	0.239
SY	**0.927**	0.361
SY3IS	**0.916**	0.198
TR	**0.853**	0.491
SY1IS	**0.844**	0.272
SY2IS	**0.820**	0.441
SY3	**0.798**	0.427
SY1	**0.765**	0.454
SY2	**0.742**	0.561
AS	**0.453**	0.238
TK	0.097	**0.873**
HR	0.344	**0.782**
TC	0.524	**0.679**
FA	0.634	**0.668**
BL	0.511	**0.597**

3.4. Calculation of factor scores and comparison of rankings

Since reducing dimensions is one of the purposes of factor analysis, factor values for each unit are calculated. With the assist of 2 factor scores obtained with the factor analysis, the alternative rankings of the teams have been checked. The 15 variables available before starting the factor analysis have been reduced to 2 factors and also factor scores equal in numbers to factor number have been obtained. 2009-2010 BEKO Basketball league ended with the rankings of the teams given in Table 3.

Table 3. 2009-2010 BEKO Basketball league team rankings

No	Teams
1	Efes Pilsen
2	Fenerbahçe Ülker
3	Banvitspor
4	Beşiktaş Cola Turka
5	Türk Telekom
6	Pınar Karşıyaka
7	Bornova Belediye
8	Erdemirspor
9	Galatasaray Cafe Crown
10	Tofaş
11	Antalya BŞB
12	Mersin BŞB
13	Aliağa Petkim
14	Oyak Renault
15	Kepez Belediye
16	Darüşşafaka Cooper Tires

After the factor analysis conducted on the statistical indicators of the match performances of the teams exhibited throughout the season, the alternative ranking as per the 1st factor, which has the greatest variance explanation rate, has been obtained as shown in Table 4.

As indicated in Table 4, in the ranking made as per the 1st factor, which has the greatest variance explanation rate, if a ranking had been made based on the performances the teams exhibited during the matches, the champion of 2009-2010 BEKO Basketball league would be Fenerbahce Ulker team, instead of the original champion Efes Pilsen. It has been also noted that the team Mersin BSB, which originally completed the league at the 12th rank, would have the 6th rank according to the evaluation based on performance data. Consequently, it will be a true approach to say that although this team has exhibited a good performance, it could not achieve adequate number of baskets. In addition the 7th ranked Bornova Belediye team would have been in the 13th ranking according to its performance evaluation. This may be the indication that the team, which exhibited a poor performance throughout the season, has achieved excessive quantities of baskets and obtained a higher ranking.

Table 4. The ranking after the factor analysis.

No	Alternative ranking	Factor scores
1	Fenerbahçe Ülker	1.91418
2	Efes Pilsen	1.85165
3	Beşiktaş Cola Turka	1.38747
4	Türk Telekom	0.65886
5	Banvitspor	0.48294
6	Mersin BŞB	-0.11165
7	Pınar Karşıyaka	-0.12523
8	Galatasaray Cafe Crown	-0.15406
9	Erdemirspor	-0.30243
10	Aliağa Petkim	-0.44048
11	Antalya BŞB	-0.46324
12	Oyak Renault	-0.62179
13	Bornova Belediye	-0.68683
14	Darüşşafaka Cooper Tires	-0.96719
15	Kepez Belediye	-1.07774
16	Tofaş	-1.34446

4. Discussion

In line with the results of the study, factor analysis as one of the multivariate statistical methods, have been conducted in order to being able to consider performance evaluations as an alternative to the evaluation of the BEKO basketball league outcomes. By scaling the statistical indicators of the performances of the teams participated in the 2009-2010 BEKO Basketball league with factor analysis, an alternative ranking for these teams has been tried to be generated. This alternative ranking is generated not by considering the number of baskets the teams scored at the end of the matches but by considering the performances of the players as a team, as well as the points they have achieved. Consequently, the ranking of the evaluation of the performances the teams exhibited throughout the season has been obtained. This has given prominence to the importance of performance.

The facts that the Kaiser-Mayer-Olkin (KMO) criteria has produced an acceptable value and that the factor loads have produced suitable values has indicated that the interpretation of the analysis results will be convenient. If the value obtained from the KMO test would be less than the value 0.5, that would have shown that the relations between pairs of variables could not be explained with other variables and in such case the factor analysis would be abandoned (Çankaya et. al., 2009). The 15 factors available at the start of the study have been reduced to 2 factors that meet the normal distribution condition and that do not have the multi-collenarity problem with the analysis. Factor scores with these features can be used in analysis such as regression, discriminant, logistics and in independent and multivariate variance analysis (Özdamar, 2010; Rencher, 2002).

In the rotated factor matrix each column vertically shows the load of each variable on the factor, while the rows horizontally show the relations of the variables with each factor (Cengiz and Kılınç, 2007). Since the factor loads of all variables are positive, it has been determined that the variables have the same direction with the other variables in the factor and that variables with same features have been gathered together.

5. Conclusion

In the study, the original rankings of 16 teams have been compared with the rankings generated with the Principal Components Analysis. The 15 variables included in the model have been considered as performance determiners and according to these variables gathered under two factors a different success rating with Principal Components Analysis have been obtained. Considering the fact that the explained variance rate has been 78.996 %, if the matches would be replayed with the same conditions, having Fenerbahce Ulker team in the 1st place instead of Efes Pilsen, having the 12th Mersin BSB team in the 6th ranking and the 7th Bornova Belediye team in the 13th ranking is considered as a high probability. The fact that the original scores and the scores predicted by PCA are not parallel to each other can be related with the 21 % unexplained variance. This study is considered important also in terms of setting forth the factors affecting the players' performances and from this point of view it is considered that the findings of the study will provide a contribution to the sports area.

6. References

[1]   Zırhlıoğlu, G and Karaca, S. 2005 genç bayanlar dünya voleybol şampiyonasına katılan sporcuların kümeleme analizi ile incelenmesi. *Hacettepe J. Of Sport Sciences*. 2006, **17**(1): 20-25.

[2]   Cengiz, D and B, Kılınç. Faktör Analizi İle 2006 Dünya Kupası'na Katılan Takımların Sıralamasının Belirlenmesi. Marmara Üniversitesi. İİBF Dergisi, Cilt:XXIII, Sayı:2, 2007.

[3]   İnce, G., & Şen, C. Adana ili'nde deplasmanlı ligde basketbol oynayan sporcuların problem çözme becerilerinin belirlenmesi. *SPORMETRE Beden Eğitimi ve Spor Bilimleri Dergisi*. 2006, pp. 4, 5-10.

[4]   Özdamar, K. Paket programlar ile istatistiksel veri analizi -2: (Çok değişkenli analizler): SPSS-MINITAB, Kaan Kitapevi, Eskişehir, (2010).

[5]   Johnson RA, ve Wichern DW. *Applied Multivariate Statistical Analysis*(6th Ed). Pearson, New Jersey, 2007.

[6]   Çokluk Ö, G Şekercioğlu and Ş Büyüköztürk. Sosyal Bilimler için çok değişkenli istatistik. SPSS ve LISREL uygulamaları. Pegem Akademi Yayıncılık, Ankara, 2010.

[7]   Jöreskog, KG and D, Sörbom. LISREL 8: Structural equation modeling with the SIMPLIS command language, Chicago, IL: Scientific Software International.

[8]   Hair, J.f., Anderson, R.E., Tatham, R.L. & Black, W.C. *Multivariate data analysis*. New Jersey: Prentice Hall, 1998.

[9]   Rencher, AC. *Methods of Multivariate Analysis*(Second Edition). John Wiley&Sons, Inc., USA, 2002

[10] Çankaya, S., Altop, A., Kul, E ve Erener, G. Faktör analizi skorları kullanılarak Karayaka kuzularında canlı ağırlık tahmini. *Anadolu Tarım Bilimleri Dergisi*. 2009, **24**(2):98-102.