NonParametric Combination method for data analytics in basketball matches

Metodo di combinazione non parametrica per l'analisi dei dati nelle partite di basket

Elena Barzizza, Nicolò Biasetton, Riccardo Ceccato, Marta Disegna and Giacomo Vezzosi

Abstract This study examines the usefulness of the NonParametric Combination technique in comparing basketball players' performance. After an extensive review of the literature on indices that can be computed to evaluate individual and team performance in basketball matches, the NonParametric combination technique is used to compare two players in the same role both on each performance index and on overall. Data collected during the 14 matches played by the basketball womens team at the University Sport Centre Padova in the 2022 winter season have been analysed in this study. Three couples of players have been analysed and the results show that only in one couple, one player was overall superior to the other.

Abstract *In questo studio viene analizzata l'utilità della tecnica di NonParametric Combination nel confrontare le prestazioni dei giocatori di basket. Dopo aver condotto un'estesa revisione della letteratura sugli indicatori che possono essere calcolati per valutare le prestazioni individuali dei giocatori e della squadra, la tecnica di NonParametric Combination è stata utilizzata per confrontare due giocatori nello stesso ruolo sia sui singoli indicatori che a livello complessivo. I dati*

Elena Barzizza
Department of Management and Engineering, University of Padova, Stradella S. Nicola, 3, Vicenza, 36100, Italy, e-mail: elena.barzizza@phd.unipd.it

Nicolò Biasetton
Department of Management and Engineering, University of Padova, Stradella S. Nicola, 3, Vicenza, 36100, Italy, e-mail: nicolo.biasetton@phd.unipd.it

Riccardo Ceccato
Department of Management and Engineering, University of Padova, Stradella S. Nicola, 3, Vicenza, 36100, Italy, e-mail: riccardo.ceccato.1@unipd.it

Marta Disegna
Department of Management and Engineering, University of Padova, Stradella S. Nicola, 3, Vicenza, 36100, Italy, e-mail: marta.disegna@unipd.it

Giacomo Vezzosi
Department of Management and Engineering, University of Padova, Stradella S. Nicola, 3, Vicenza, 36100, Italy, e-mail: giacomo.vezzosi.00@gmail.com

raccolti durante le 14 partite giocate dalla squadra femminile del Centro Universitario Sportivo di Padova nella stagione invernale del 2022 sono stati analizzati in questo lavoro. Tre coppie di giocatrici sono state analizzate e i risultati dimostrano che solo in una coppia una giocatrice è globalmente superiore rispetto all'altra.

Key words: NonParametric Combination, Permutation test, Basketball

1 Introduction

In order to perform comparisons between pairs of players, the NonParametric Combination (NPC) technique [1] can be applied. NPC is a highly flexible permutation-based methodology which allows us to deal with complex problems, and, in the context of basketball data, it can be used to identify significant differences, in terms of multiple performance indicators, between players in the same role.

Let us assume that we want to compare the performance of two players, named P_1 and P_2, to verify whether one player outperforms the other. For each player we have information on V variables (i.e. performance indicators). The system of hypotheses of interest can be written as follows:

$$\begin{cases} H_0 : \mathbf{M}_1 = \mathbf{M}_2 \\ H_1 : \mathbf{M}_1 > \mathbf{M}_2 \end{cases} \tag{1}$$

where $\mathbf{M}_k$, $k = 1,2$, is a vector of V statistics (such as median or mean) computed for the k-th player. The NPC methodology initially decomposes this system of hypotheses into V sub-systems (i.e., one for each performance indicator). The generic v-th sub-system of hypotheses ($v = 1,\ldots,V$) is as follows:

$$\begin{cases} H_{0v} : M_{1v} = M_{2v} \\ H_{1v} : M_{1v} > M_{2v} \end{cases} \tag{2}$$

After choosing an appropriate test statistic, NonParametric Combination addresses each sub-system individually and provides us with V partial p-values achieved through permutation, so that we can detect differences between players in terms of each performance indicator. It should be noted that the same permutation mechanism is applied for the computation of each partial p-value [1], so that we implicitly take into account the existing dependence between variables. Then, a combination step in the NPC algorithm allows us to retrieve a combined p-value, which merges the insights provided by the partial p-values, to be used to address the multivariate problem of interest expressed in the system of hypotheses (1).

This methodology only requires us to make two fundamental choices: the choice of the test statistic and the choice of the combining function to be adopted in the final combination step. The first choice is driven by the nature of the problem at hand. Since in this study we deal with numerical variables (see Section 2) characterised

by outliers, the multivariate medians have been used for comparison between players. Therefore, the test statistic related to the v-th performance indicator is simply computed as a difference between medians as follows: $\text{median}_{1v} - \text{median}_{2v}$

The second choice is mainly driven by the number of available sub-problems in which the null-hypothesis is expected to be rejected and the correlation between variables [9]. In this study, we rely on Fisher's combining function, as it has been shown to be the best solution in many scenarios.

2 Data description

The case study analysed in this research refers to 14 matches played by women's basketball team of the University Sport Centre (CUS) Padova during the 2022 winter season. The official roster is made up of 16 players. Given a specific game, for each player who took part in the game, several information have been collected, including: shooting statistics (points scored; field goals attempted and made, separating also 2 and 3-point shots; free throws attempted and made; throws attempted and made from each part of the pitch), assists, rebounds, blocks, steals, fouls, turnovers and minutes played.

2.1 Basketball performance indexes

As described in [4] and [3], several indicators have been computed to evaluate the performance of players. Based on the data collected during the matches, the following indeces have been computed:

- Performance Index Rating (PIR) [6]. This metric allows us to assess the efficiency of players in a match, summing up all the "positive" actions (i.e. points, total rebounds, assists, steals, blocks, received fouls) and subtracting from this quantity all the "negative" actions (i.e. missed field goals, missed free throws, fouls committed, turnovers). The disadvantage of this index is that it does not consider neither the importance of the individual statistic among all the statistics nor the role of the player in the match.
- Player Impact Estimate (PIE) [7]. This index is computed for each player, for each team, and for each match. It allows us to obtain a measure of the overall contribution of the player and the team to each match. This index is useful for comparing players and teams.
- Floor Impact Counter (FIC) [5]. This index allows us to obtain a player's evaluation rating system in which greater importance is given to assists, offensive rebounds, and to the construction of offensive actions in general.
- Offensive Efficiency (OE) [8]. This index is used to measure the quality of player's offensive production. The index is computed as the rate between the

number of profitable offensive possessions in which the player has been involved and the total number of potential end-of-possession situations for the player.
- Efficient Offensive Production (EOP) [8]. This index is derived from the previous one and is computed as: (0.75 $\times$ assists + points) $\times$ OE. The tuning parameter of the assists (equals to 0.75) represents the estimated contribution of an assist to the final points scored.
- Adjusted Field Goal (AFG) [6]. This index allows us to obtain a measure of the shooting ability of each player.

2.2 Data analysis

The performance indicators computed as described in Subsection 2.1 have been divided by the minutes played by each player in each match before to perform the descriptive and inferential analysis. The analysis have been conducted on 6 players of the same team. The descriptive analysis allowed us to identify potential differences between pairs of players in the same role, comparable in terms of played matches and minutes played (see Figure 1). In particular, the pair of players is formed as follows: group 1 is made up of player 1 and player 2; group 2 is made up of player 3 and player 4; group 3 is made up of player 5 and player 6. Looking at Figure 1 it is possible to notice that in the first group player 2 outperforms, in terms of the median value for the EOP and the OE indices, the teammate. In the second group, player 3 outperforms (in terms of median value) player 4 with respect to FIC, PIE and PIR indices. Finally, in the third group, player 5 appears to perform better looking at the median value of the PIE index with respect to player 6.

3 Findings and conclusion

The results of the permutation-based testing procedure are reported in Table 1. Please note that the problem of multiplicity control arises when the number of sub-hypotheses to be tested is greater than one. For this reason, we applied the Bonferroni-Holm method to adjust partial p-values as indicated in [1].

As anticipated from the descriptive analysis, player 1 significantly outperforms player 2 in terms of EOP, with an adjusted partial p-value equal to 0.026. Regarding group 2, player 3 outperforms player 4 with respect to the PIR, PIE, and FIC indices (p-values are all lower than 0.05). Finally, in group 3, player 5 outperforms player 6 in terms of PIE only (p-value 0.044).

The combined p-values give us further useful indications regarding the players' overall performance. Using this combined test, it is possible to affirm that overall player 3 outperforms player 4 (combined p-value equals 0.006) while in group 1 and group 2 it is not possible to identify a player who overall outperforms the other

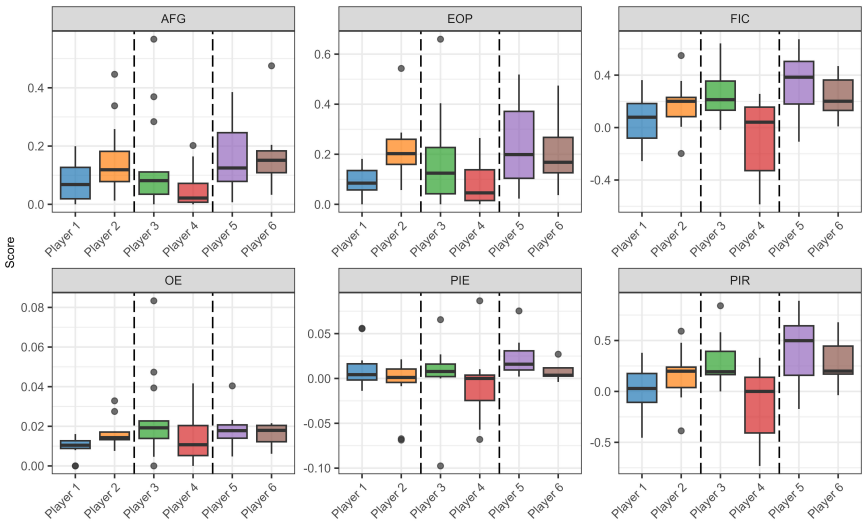

Fig. 1 Distribution of the different performance indicators per player.

Comparison	PIR	PIE	FIC	OE	EOP	AFG
$P_1 < P_2$	0.318	0.791	0.277	0.318	0.026	0.430
$P_3 > P_4$	0.014	0.014	0.014	0.134	0.544	0.669
$P_5 > P_6$	0.582	0.044	0.546	0.844	0.844	0.776

Table 1 Partial p-values.

one (the combined p-value are equal to 0.051 and 0.197 respectively for group 1 and group 2).

The proposed permutation-based test can be easily extended to the comparison of more than two players, integrating it with the multivariate ranking procedure by [2].

Acknowledgements This study was carried out within the MICS (Made in Italy Circular and Sustainable) Extended Partnership and received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 D.D. 1551.11-10-2022, PE00000004). This manuscript reflects only the authors views and opinions, neither the European Union nor the European Commission can be considered responsible for them. This study has also been supported by the following projects: BIRD 2022 project titled Fuzzy theory in Unsupervised Machine Learning algorithm and Sentiment Analysis; PNRR MOST - Centro Nazionale per la MObilità soSTenibile, Spoke8: Maas & Innovative services.

References

1. Pesarin, F., Salmaso, L.: Permutation tests for complex data: theory, applications and software. John Wiley & Sons (2010)
2. Arboretti, R., Bonnini, S., Corain, L., Salmaso, L.: A permutation approach for ranking of multivariate populations. Journal of Multivariate Analysis. 132, 39–57 (2014)
3. Metulini, R., Gnecco, G.: Measuring players importance in basketball using the generalized Shapley value. Annals of Operations Research. 1–25 (2022). https://doi.org/10.1007/s10479-022-04653-z
4. Corain, L., Arboretti, R., Ceccato, R., Ronchi, F., Salmaso, L.: Testing and ranking on round-robin design for data sport analytics with application to basketball. Statistical Modelling. 19(1), 5–27 (2019)
5. Ferrario, A.: Basketball Analytics: the use of data science to describe and predict the performance of an NBA team. *Master's thesis, Università degli studi di Milano Bicocca*, (2021)
6. Cene, E., Parim, C., Özkan, B.: Comparing the performance of basketball players with decision trees and TOPSIS. Data Science and Applications. 1(1), 21–28 (2018)
7. Senatore, J.V., Fellingham, G., Lamas, L.: Efficiency and productivity evaluation of basketball players' performance. Motriz: Revista de Educação Física. 28 (2022)
8. Lee, D.-J., Page, G.L.: Big Data in Sports: Predictive Models for Basketball Player's Performance. Mathematics in Industry Reports. (2021)
9. Langthaler, P.B., Ceccato, R., Salmaso, L., Arboretti, R., Bathke, A.C.: Permutation testing for thick data when the number of variables is much greater than the sample size: recent developments and some recommendations. Computational Statistics. 38, 101–132 (2022)