

Statistics in Sports

Project BDSports (Big Data Analytics in Sports)
bodai.unibs.it/BDSports

Paola Zuccolotto^a, Marica Manisera^a, and Ron Kenett^b

^a*Big & Open Data Innovation Laboratory (BODaI-Lab), University of Brescia (Italy)*

^b*KPA Ltd and Neaman Institute, Technion (Israel) - University of Turin (Italy)*

November 15, 2017

This Special Issue has been promoted by BDSports (Big Data Analytics in Sports), a project developed as part of the activities of the Big & Open Data Innovation Laboratory of the University of Brescia, born in 2016 thanks to the financial support of Fondazione Cariplo and Regione Lombardia.

BDSports is designed to set up a unique collaboration of experts interested in sport analytics both from a scientific and a practical point of view. The goal is to create a network able to facilitate contacts and joint research initiatives. Specifically, the project aims to organise events, carry out Special Issues in scientific journals, share ideas and data in order to publish scientific and non-scientific papers, collaborate with teams in various sports by supplying them analytics and apply for research grants. The data scientists expertise covers a wide range of quantitative tools in the fields of statistical modelling, multivariate data analysis, data mining, algorithmic modelling and machine learning.

Editorial

In the last decades the idea of a statistical thinking in sports has gained a rapidly growing interest, as documented by the wide scientific production on this theme. This Special Issue collects qualified research papers that propose developments of new statistical methods of sport analytics and give new insights into sports.

The papers published in this Special Issue focus on basketball, soccer, tennis, American football and cricket (4, 4, 2, 1 and 1 papers, respectively) and cover a wide range of topics, such as technical issues concerning the game and the way of playing, performance analysis, sport management and betting.

More specifically, among the papers dealing with basketball, players' scoring efficiency is considered by Antonio Pulgarín García, José Pablo Arias Nicolás and Hector Valentín

Jimenez Naranjo, who propose a Bayesian approach considering both individual scoring skills and collective performance, and by Manuel Ruiz Marin, Fernando Lopez, Jose Antonio Martinez and Martí Casals, who pay special attention to stability and regularity of scored points. Rodolfo Metulini deals with the visualization of data recorded by means of tracking systems that produce spatio-temporal traces of player trajectories with high definition and frequency. Finally, Federico Bianchi, Tullio Facchinetti and Paola Zuccolotto propose a data mining procedure based on fuzzy clustering and self-organising maps to analyse players' performance in order to define new roles overcoming the old classification into the traditional five positions.

Moving to soccer, the number of goals scored in a match is modelled with a time series analysis approach by Naushad Mamode Khan, Yuvraj Sunecher and Vandna Jowaheer, who propose the use of a bivariate integer-valued autoregressive process with parameters estimated by a GQL algorithm based on higher ordered covariances, while Emilio Gómez-Déniz and Nancy Dávila-Cárdenes propose the use of a compound Poisson distribution to identify the factors affecting the number of yellow and red cards shown during a competition. Sport management and financial issues are taken into account by Yuri Zelenkov and Ilya Solntsev who develop a methodology for quantitative analyses of clubs' efficiency, while Fabio Fin, Maria Iannario, Domenico Piccolo and Rosaria Simone analyse experts' and laymen's perceptions about players' performance and the resulting players' rankings, with a specific focus on the intrinsic uncertainty of decision processes.

The two papers focused on tennis deal with the assumption of *iid*-ness of scored points and betting, respectively. More specifically, Andrea Carrari, Marco Ferrante and Giovanni Fonseca postulate that in any game different winning probabilities correspond to some selected game situations and analyse this assertion by means of a Markovian model. Francesco Lisi and Germano Zanella propose a logistic regression model to predict the winning probability in a tennis match and investigate whether it is able to beat bookmakers predictions.

In the paper on NCAA football, Eric Howington and Nathan Moates investigate on whether an effective advantage derives from the extra time for preparation, rest and recovery due to a bye week, and conclude that the off-week advantage is actually a "myth".

Last but not least, the paper on cricket considers the so-called "rain rules" used to decide the outcome of a game in case of a weather interruption. With the aim of developing a fair rule, Jack Hogan, Edward Cohen and Niall Adams construct a model and empirically compare its performance to that of traditional models proposed in the literature.

It was a pleasure for us to have the opportunity to read so many interesting studies, and we wish to express our thanks to all the authors and reviewers that have made this Special Issue possible. We also thank the EJASA editors (Amjad D. Al-Nasser - Editor in Chief, Maurizio Carpita - Co-Editor in Chief, and Enrico Ciavolino - Executive Managing Editor) for supporting us during the whole publication process.

Paola Zuccolotto, Marica Manisera, Ron Kenett (Guest Editors of EJASA)