

# Sports Analytics

## When Statistics and Data Analytics Meet Sports

By

Dimitris Karlis, Professor, Department of Statistics, AUEB

Ioannis Ntzoufras, Professor, Department of Statistics, AUEB

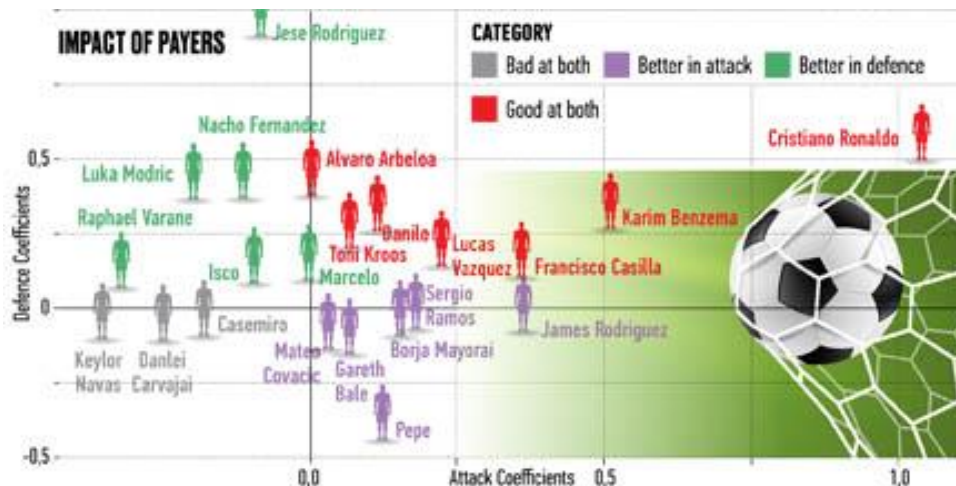
Sports has always been a fertile ground for the application of statistical methods, over a wide range of problems of investigation. For instance, one may be interested in predicting matches' final outcomes, measuring and/or visualizing the performance of a particular team or its individual players, and/or evaluating the impact of athletes' and strategies' performance throughout the match (i.e. when to proceed with player substitutions, or which, the best-performing basketball team, was). Such decisions, although of main attentiveness to those directly involved in a specific sport of interest, may as well incite the interest of fans, students or even the general public.

"Moneyball", a 2011 American biographical sports drama film, describes the successful 2002 season of the Oakland Athletics baseball team and the efforts of their general manager (Billy Beane) to assemble a competitive team, based on the empirical analysis of baseball statistics measuring activity during play. This very movie, based on a true story, popularized the idea of using data as a means of decision-making in sports.



Nowadays, especially in major leagues, teams of expert scientists scrutinize data, to reveal the opponent's tactics.

Today's need for larger, in quantity, and better, in quality, data in sports has increased the use of state-of-the-art data collection techniques. These days, teams make use of technologies capable of attaining detailed – real-time – player physiological measurements, according to which, the physical condition of the latter may afterwards be determined, or even, their next moves on the field of play may be predicted. This results in the constant feeding of sports scientists with more data, hence creating a cycle of data demand and use. Certainly, neither are all efforts successful, nor are they always accurate, but the increase in the availability of data does, ultimately, indeed, create the path towards new methodologies and scientific challenges, in the entire sports industry.



Within this context, data-based sciences, such as Statistics, come to the fore, allowing for numerous options with regard to solving problems and questions related to/arising from sports. Taking advantage of modern technology as well as the abundance of available data, these sciences may, therefore, help us understand and predict the outcome of sporting events, while simultaneously being utilized as a means to design fair and competitive sporting events, significantly increasing the quality of the produced entertainment product, which in turn may lead to increased ticket demand and television/broadcasting rights. For instance, in the World Cup draw, teams create different seeding groups based on an algorithm determining their capacity, so that teams are being distributed more fairly and so that greater competitiveness is achieved, thus, stimulating greater interest in all end users of the product.

Predicting the outcome of a match or sporting event is easy in every way. However, the uncertainty of the final result might be processed through statistical methods, so that better forecasts are obtained. By assessing historical data, one may, not only understand how to better design leagues and sporting events that will increase competition, but also, how to measure player performance, and reveal ways in which a coach could improve their team's performance or successfully tackle an opponent. Mathematics, Statistics and machine learning can provide major assistance to all members of the sports industry (teams, coaches, players, consumers, and fans) concerning the improvement of their performance or in making optimal decisions.

Indubitably, we must not forget the "entertainment" aspect of sports. Sports fans attend sporting events for the excitement that they provide. Guessing the final result of a game is of great excitement/challenge for fans and so is the uncertainty, surprise and unpredictability that accompanies it. Sports characterized by less uncertainty tend to be less exciting. The generated feeling when a team scores a last-minute goal is, perhaps, hard to be found in other real-life activities. Hence, understanding and figuring out ways to support/promote this uncertainty is considered to be the "hidden gem" of sports. Statistical science helps to understand and measure this uncertainty and, although in most sports, predicting the winner may not be possible, it

certainly provides an improved “sports experience”, by highlighting evidence, which, up until recently, was unavailable.

**AUEB Sports Analytics Group:** The **AUEB Sports Analytics Group** has been operating since 2015, within the Department of Statistics and the Computational and Bayesian Methodology Laboratory, with a multitude of activities (theses, research publications, publications in the local press), whilst it also runs the AUEB Sports Analytics Workshop, every year, with scientists, distinguished in the field (both from Greece and abroad), attending the event. <https://aueb-analytics.wixsite.com/sports>