APPLICATION OF STATISTICAL INDICATORS IN FOOTBALL

Margit CSIPKÉS, György CZIMBALINECZ
University of Debrecen Faculty of Economics Institute of Statistics and Methodology, Debrecen, Hungary
csipkes.margit@econ.unideb.hu
czimbagyuri@gmail.com

Abstract: *In our article, we describe the different statistical indicators using secondary research. In our article, we show you in which areas of football the data can be used for analysis. We present in detail the data that can be used for player monitoring and sports betting. With the help of the primary research, we created an online questionnaire in which we asked footballers, professionals and fans about the topic. The results obtained are illustrated with figures and tables.*
It is clear from the research that professionals, footballers and supporters share the same view, as the economic situation caused by the epidemic means that club teams have to rely much more on data-driven player monitoring. It's a good idea to invest in players instead of buying overvalued football players. As described in our article, it is not possible to predict the outcome of a match with traditional statistical indicators (goal shots, corners, possession of the ball, etc.). This is because these statistics do not return an accurate picture of a match. These numbers are not suitable for drawing conclusions for a match. In our opinion, these numbers show the quantity and not the quality of the shot on goal.

Keywords: *sports, indicators, calculation, football analysis*

JEL Classification: L83, Z20, Z21, Z28

1. Heading: Generally, Introduction or Background- Importance of statistical indicators in football

In football, one of the most significant statistical indicators is the Expected goals ratio (in the literature it is the xG indicator), which determines the proportion of goals expected. This is a statistical model that shows how likely an average footballer is to score a goal in a given match. There are many models to calculate that can take into account different variables and conditions. The most significant variables are:
- the exact location of the shot: the farther or the narrower the angle of the shot, the less chance the player has of scoring;
- the body part: with which part of the body the player scored the goal (it is more difficult to control the ball upside down than from the ground);
- type of goal pass: the chance to score a goal changes if the player making the shot wants to score a goal after a bounce, a pass or a point kick

- types of attacks: a lot also depends on this variable (multiplayer, counterattack, series or a simple cut-off, etc.).

Based on the variables, the scoring attempt is compared to thousands of similar shots, the end result of which is assigned a number between 0 and 1 for that shot (1 is the safe goal and 0 is the worthless option). For example, if a shot gets a value of "0.5 xG," the sample has a 50% chance of scoring. The value of all the shots of the teams in a given match is added up, so by the end of the 90th minute, it will be decided which team has developed better in terms of the number and quality of shots fired.

Figure 1 shows an xG diagram in which the goals are marked with the green circles and the omitted positions with the red circles.
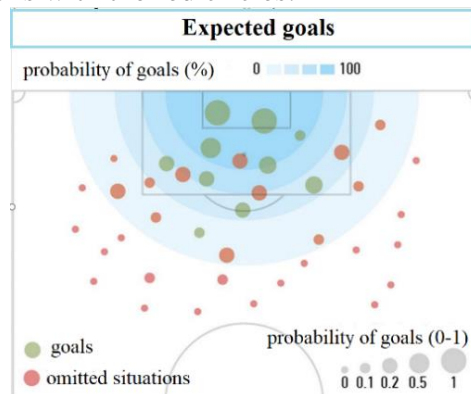
Figure 1: Probability of expected goals and their scoring
Source: nemzetisport.hu

The larger the circle in the figure, the more likely it is to score a goal. The minimum value that can be assigned to a shot is zero, and the highest value is 1 (the probability of this is very minimal). Based on this, if a shot is worth "0.5 xG," that shot will have a 50% chance of scoring. Blue semicircular curves are used to assign the value to the shot. The farther away and the more displaced a player is, the less likely he is to score a goal.

The other most significant indicator is the Expected assists ratio (known in the literature as the xA indicator), which is an indicator of expected goal pass. In a goal pass, only the pass from which the teammate scores a goal counts, so a player can create up to five positions in a match if his teammates do not change to a goal. The amount of key passes gives a more accurate picture, as it already takes into account every pass that is followed by a shot. However, with the help of the xA pointer, we can find out how much the player who gave the last pass before the shot put his teammate in position. This indicator also follows the principle of the xG model, with each pass before the shot being assigned an xG of 0 to 1 for the subsequent shot. The higher the quality of the situation based on the xG score, the more the preparatory player puts his teammate in the xA score. A pass in preparation for a shot from thirty meters in the middle of the opponent's half will be worth much less than a pass after which the teammate only has to throw the ball into the goal.

The third most significant indicator is the Expected goal against (xGa indicator in the literature), which gives the number of goals expected to be scored. The lower the number of this indicator, the better the defense of that team. By calculating the difference between the xG index and the xGa index and comparing it with the real goal difference, we can find out whether the team under test performed according to its performance during the season or just below or above.

The non-Penalty Expected goals (NPxG index in the literature) is also significant, giving the proportion of goals expected without a penalty. With Xg, we can not only measure the performance of teams, we can also measure the performance of players. Comparing an attacker's actual number of goals to the expected number of goals, we can get an idea of how successful a given player is in a position, how good a finisher is, or just how unlucky he is in front of the goal. What can distort judges 'performance is the performance of penalties, as it has a penalty of" 0.76 xG, "which can thoroughly discard the total xG score for a designated penalty executive. That's why it's worth subtracting these from the formula and looking at the so-called NPxG, which takes the shots fired from the penalties from the total value: total xG - penalties xG = NPxG value

Post shot expected goals (PSxG) examines the quality of the shot. This means the chances of scoring from that experiment. If a shot doesn't find a goal, there's no chance of it scoring. In this case, the value of PSxG will be around 0. For shots that hit a goal, the closer and farther a shot goes to one edge of the goal, the higher it will receive (for example, a shot from twenty yards to the center of the goal will be less than a shot from eight meters to a corner). The most advantageous feature of the PSxG is that it is easier to measure the performance of the goalkeepers. The number of goals scored without goals or the effectiveness of defense alone doesn't say much about how well someone defends. If a team lets a few shots into its goal (and they also come mostly from a distance), the goalkeeper has a much better chance of getting off the goal without the goal being scored (while the goalkeeper is constantly blocking weak attempts). The PSxG shows exactly the quality of shots that a particular goalkeeper must deliver.

Passes per defensive action (PPDA) means passes allowed before intervention. This indicator shows how many passes a given team has allowed the opponent to play in their own half of the field before carrying out any defensive action (ball acquisition, preventive assembly, or foul). The lower this score, the more intense the attack on that band. It is also important to point out that the PPDA indicator strictly measures the intensity of pressing and not its quality. It can easily happen that a team allows very few passes to an opponent on average (trying hard to get the ball), yet they create high-quality situations against them due to unsuccessful attempts to attack. The PPDA indicator gives a bad result if a team allows the opponent to pass in the back (increasing the number of passes before the defensive action) but starts very aggressively on a signal (such as a ball next to a line, a bad pass, a back pass) to attack.

The Opponent passes allowed per defensive actions (OPPDA) indicator determines how many passes a team is able to make in its own half of the space before the

opponent takes a defensive action. Unlike the PPDA indicator, the higher the value of the OPPDA indicator for a given team, the better it can withstand the opponent's attacks. However, this is often only true on paper. If a team is left free to pass in the forefront of their own sixteen and then the player regularly loses the ball in an aggressive attack launched on a signal, no matter how high the OPPDA score is, it will not be more resistant to pressing. Of course, it is also true that teams that are able to replay aggressive pressing on a regular basis after a while may no longer be caught attacking by opponents, which can result in an even higher OPPDA score (I01).

The Expected goal ratio chain (xGChain) indicator shows how much each player takes part in the build-up play. This indicator is derived from the xG model. When determining the xGChain value, each ball possession cycle is recorded. The shots fired during these are recorded, and the corresponding xG values are added up. The player who touched the ball during the action is then assigned to the player.

The Expected goal 2 / Expected save (xG2 / xSV) metrics show which goalkeepers are used to measure the ratio of expected defenses. This value shows, from the perspective of the goalkeepers, the probability that a goal will usually be scored from a given shot. Only shots that hit the goal are counted, and their value depends on the strength of the shot fired. The xG indicator indicates its danger at the moment before the shot, and the xG2 indicates the moment after the shot. A position in front of a goal has a distinctly high xG value, but if the player does not hit the goal from the shot, the value of xG2 is already 0.

If we subtract the value of xG2, we immediately get the expected save index, which shows $(1 - xG2 = xSV)$. The indicator determines how likely the goalkeeper is to fire a given shot.

The Indice di Pericolosita (IPO) indicator is known as the hazard indicator. This indicator was introduced in Italy, an improved (more reliable) version of the xG indicator. The indicator measures the degree of danger of a team's attack leads, which is calculated on the basis of the total score of the goal situations (promising attack leads, shots, corners, passes, key passes) achieved in the attacking third. Statisticians evaluate the performance of the bands based on a constant table. The cleaner the situation, the more points a team puts at the end of their opponent's goal (the better their chances of scoring a goal). Consider an example: completing a penalty (or if the attacker can lead the ball alone) 10 points, while taking a free kick from 20 meters (a shot taken within a penalty area, a pass or a corner) can score 4, 3, 2 or 1 point.

The Indice di Rischio Difensivo (IRD) is abbreviated to the risk indicator. It was introduced in Italy, which measures the degree of passive danger of ensembles (how much space the ensemble gives to the opponent's attacks). The performance of the teams is assessed on the basis of similar criteria and a constant table as for the hazard indicator. Subtracting the results of the two metrics (IPO-IRD) gives the extent to which the team under study dominated the match or just played a subordinate role (I02).

Team	IPO	IRD	IPO - IRD	Team	IPO	IRD	IPO - IRD
Napoli	61.8	25.4	36.5	Sampdoria	44	47	-3
Milan	66.7	35.1	31.6	Lazio	40.3	45.8	-5.6
Atalanta	57.3	38	19.3	Torino	38.2	45.4	-7.2
Udinese	46.7	27.6	19.1	Cagliari	40	49.4	-9.4
Inter	57.2	38.3	19	Verona	43.5	53.4	-9.9
Roma	60.1	45	15.1	Crotone	34.3	51.3	-17
Juventus	52.4	37.9	14.5	Parma	33	53.6	-20.6
Sassuolo	61.1	50.8	10.3	Spezia	38.1	62.9	-24.8
Bologna	56.2	53.6	2.6	Genova	27.1	53.6	-26.5
Fiorentina	48.9	49.4	-0.5	Benevento	38.1	74.6	-36.4

Table 1: The IPO-IRD table for the 2020-21. until the 7th round of the Italian First Division
Source: Own editing based on bunteto.com

In addition to the names of the teams, the points of the IPO show how much the team generally posed a threat to their opponents during the matches (Table 1). The higher this score, the more dangerous the team. These are marked with different colors in the table: green for outstanding performance, blue for good performance, yellow for poor performance, red for poor performance. The IRD column shows how vulnerable (how vulnerable) the troops' defenses are. In this column, the lower the score, the better the defense, the higher the score, the higher the risk. In the IPO-IRD column, the difference between the two indicators shows how much the team is usually dominated by the team or how subordinate it is.

2. Research, results - What are statistics used for in football?

There are four immediate areas of application of the information processed, which are as follows:
1. One group is pre-match preparation. The data stored by the statistical analysis companies can be retrieved at any time for the bands in the partnership agreement. Abroad, it is common for video analysts to compile a material for the coach based on this data, who will use it to shape the team's preparation for the next opponent. An excellent example of this is the example of István Kovách (leader of the Central European region of Instat Football) in his football podcast called Ziccer, when a goalkeeper from Hungary measured the opponent's free kicks, corners and coached the goalkeeper for the match based on Instat's data.
2. The second group is the post-match analysis. After the match, the associations in the partnership agreement will receive a multi-page pdf document from the analysis companies, in which all the data about the players (running volume, heat map, won / lost matches) can be found. Based on these, the coach has the opportunity to analyze a situation with the given player, which may later affect the development of the player.

3. The third group belongs to the commentators. In essence, statistical firms can provide live help even to commentators who can use them immediately during broadcasts (e.g., Opta, tweet, posted curiosities, etc.).

Next is the player observation. Certain statistics allow analysts to list arbitrary teams or players based on special filtering methods. If there are less listed teams or players on these lists, it means that they will compete with the best based on certain criteria. In this case, it is worth paying close attention to them, as there is a chance that they will be able to find players with high ability and high potential for little money (I02). Brentford FC is not a big name club. This West London team has spent much of its history in the lower classes of English football. In recent years, however, they have gained a huge reputation in the transfer market. The transfer policy of the Bees (nicknamed Brentford FC) laid the foundation for them, as they also finished in the first half of the table for five consecutive years after promotion from the third division. This result was achieved by having the club with the fourth lowest budget in the Twenty-Four Team Championship for years. Brentford's successful player selection policy and sales policy have allowed them to now have a cutting-edge budget and fight for promotion to the first division. After the promotion in 2014, the club reformed its selection system. Using mathematical models, they discovered for themselves several key players that were relatively inexpensive to obtain. The success of this model is now clear. This strategy is also followed by the Danish FC Midtjylland, which is not surprisingly owned by the same person as Brentford (Matthew Benham). The English owner also has a sports betting company that uses various statistical methods to estimate the outcome of matches and the performance of teams and players. These methods have brought them innovation in transfers. On a small budget, they began to analyze the performance of players less known to others. Midtjylland President Rasmus Ankersen agrees with Benham that they should follow a data-based player policy. According to Ankersen, their model can be applied in any country, but it requires completely different thinking from management than usual, especially if they want to compete with larger clubs (EU, 2021).

The first column of Table 2 and Table 3 contains the name of the player, then the transfer details of the incoming players (the year of approval, which club they came from and the amount of the transfer fee paid by the team), then the details of the outgoing player (in which year they were sold, to which club he left and how much the team received for it) and the last line shows the profit from the sale of the player.

Player name	Year	Club	Transfer fee	Year	Club	Transfer fee	Profit
Ollie Watkins	2017	Exeter City	7 millions	2020	A. Villa	30.8 millions	23.8 millions
Said Benrahma	2018	OGC Nizza	1.7 millions	2021	West Ham	23 millions	21.3 millions
Neal Maupay	2017	St-Étienne	2 millions	2019	Brighton	22.22 millions	20.22 millions
Chris Mepham	2016	Brendford B	for free	2019	Bournemouth	13.6 millions	13.6 millions
André Gray	2014	Luton Town	620 thousend	2015	Burnley	12.4 millions	11.78 millions
Ezri Konsa	2018	Charlton	2.85 millions	2019	A. Villa	13.3 millions	10.45 millions
Scott Hogan	2017	Rochdale	950 thousend	2017	A. Villa	10.5 millions	9.55 millions

Table 2: Brentford's highest-profit player sales
Source: Own editing based on transfermarkt.de

Player name	Year	Club	Transfer fee	Year	Club	Transfer fee	Profit
Alexander Sörloth	2017	Groningen	450 thousend	2018	C. Palace	9 millions	8.55 millions
Boubacarr Sanneh	2018	Horsens	200 thousend	2018	Anderlecht	8 millions	7.8 millions
Pione Sisto	2013	Midtjylland U19	for free	2016	Celta Vigo	6 millions	6 millions
Paul Onuachu	2015	Midtjylland U19	for free	2019	Genk	6 millions	6 millions
Rasmus Kristensen	2016	Midtjylland U19	for free	2018	Ajax	5.5 millions	5.5 millions
Andreas Poulsen	2017	Midtjylland U19	for free	2018	Mönchengladbach	4.5 millions	4.5 millions
Mikkel Duelund	2015	Midtjylland U19	for free	2018	Dinamo Kijev	4 millions	4 millions

Table 3: FC Midtjylland's highest-selling footballers
Source: Own editing based on transfermarkt.de

Next up are sports betting. Football is a random game, as there are very few goals, so random plays a bigger role in the game (e.g. a bounce shot). This means that the best teams win less than in a high-scoring sport (e.g. basketball). Because we can extract statistics from a large sample, 10, 20, or even 40 matches are not enough to filter out the role of randomness in these. A professional sports bettor will create equations that show patterns based on which he will place his bets. It doesn't look at where a team is in the leaderboard, but looks for underground performance metrics that can be used to make a team's future performance likely. Sports bettors like to look at the goal difference, which is the difference between the goals scored and scored by the team, which is a simple but reliable figure.

The other favorite's favorite figure is the difference in shots, the difference between a team's shots and its shots on goal. This is also a reliable figure for sports bettors, as good teams shoot more often than bad teams.

Consider an example (Table 4). Newcastle finished fifth with a negative shot. His opponents try an average of 1.4 more shots in a match than the Magpies (nicknamed Newcastle). This is an extra high scoring efficiency. A positive goal difference is accompanied by a negative shot score (this is also an unsustainable efficiency score). The first column shows the position in the table. Then the name of the team in that position. H is the difference in the shot of the matches of the home court. V away shooting differences. T is the difference for the entire season. Calculation: The difference between the shots fired at home and away matches is added up and then divided by two [(H+V)/2 = T] (I03).

Place	Team	H	V	T
1	Man City	12.3	5.3	8.8
2	Man Utd	8.5	-1.2	3.6
3	Arsenal	9.5	3.1	6.3
4	Tottenham	10.4	2.1	6.2
5	Newcastle	3.5	-6.3	-1.4
6	Chelsea	8.8	2.8	5.8

Table 4: Shot edge differences in the 2011/12 season
Source: Own editing based on 21stclub.com

4. Research, results - Use of statistics used in football

In the survey, 2 questionnaires were completed. The first questionnaire was completed by 176 people working in the world of football. The composition of this group is 36% female and 64% male. Fillers can be divided into four age groups based on their age. The oldest among the respondents was 74 years old. The youngest was 14 years old. The respondents included a fan (30%), a footballer (32%), a coach (19%) and a sports manager (19%).
The awareness of the indicators presented earlier was measured with the first professional question (Figure 2). In addition to the expected goals (17.67%) and expected goals (16.73%), the majority of the respondents know the expected goals (14.06%) and expected defenses (13.39%), which is not surprising. their application and use is widespread. The number of passes allowed before the defensive action, the number of passes taken before the defensive action, the hazard and risk indicator are less well known, while only 5.22% of the relative frequency do not know any of the statistical indicators listed.
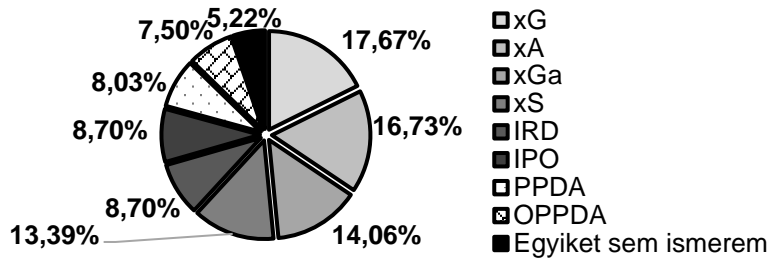

Figure 2: Knowledge of statistical indicators in football
Source: Own editing

It was important for us to look at whether the plethora of statistics and series of numbers are good for football. Two-thirds of respondents to the questionnaire believe that it has a good impact on modern football because conclusions can be drawn from such data. From these, you can find out how much a team can dominate a match or just play a subordinate role. A minority of respondents say they shouldn't deal so much with the data because on the court, the players decide the outcome of the match and not the numbers.
Since the commentator tells data about the players during each broadcast, we also looked at how much it disturbs viewers during broadcasts if the commentator keeps posting data. It can be stated here that almost half of the respondents (48%) are not bothered if the commentator says with data. These respondents believe this is also the job of the announcer. 35 percent of respondents are interested in the data that makes a match more interesting to them. 17 percent of those surveyed are disturbed by the data made during the match (Figure 18).
As half of the respondents are not bothered by the presentation of the data, we asked this group if they would also use different analysis software. Respondents are best acquainted with the InStat Football analysis software, followed by OptaSport, WyScout, and the Global Soccer Netwok (GSN-Index) (Table 5).

Softver name	Share (%)
InStat Football	30,38%
OptaSport	22,78%
WyScout	19,49%
GSN-Index	14,94%
She/He knows none of them	12,41%
In all	100,00%

Table 5: Knowledge of analytics software
Source: Own editing

Because accurate data is important when watching a match, we examined whether respondents would use a football statistics application that provides accurate data for the minute. More than seventy percent of those surveyed answered yes (34.09% of which would even use such an app as a coach).

According to the respondents, the statistical data can be used by the respondents primarily for post-match analyzes / evaluations. This was followed by player monitoring and transfer policy. Few think the team could benefit from this in pre-match preparation (Table 6).

Ranking	Area of use
1.	Post - match analysis
2.	Player monitoring
3.	Transfer policy
4.	Preparation before the match
5.	Player development

Table 6: Ranking of respondents in terms of data application
Source: Own editing

In the following, we compared the three major statistical indicators and asked the fillers which indicators they were most interested in about a match. It turned out that people are most interested in the data used for player evaluation (number of runs, number of successful tricks, etc.) and traditional statistics (goal shots, corners, etc.). These two groups were followed by the newly released Xg index et al.

As almost 80% of the respondents also used to participate in sports betting, we also examined this topic in more detail. The group of sports bettors can be divided into 2 parts: one part bets regularly (80%) and the other only occasionally bets (20%). Based on the answers, the respondents most often decide on the outcome of the bet based on the form. This is followed by inference based on bets and in-depth analysis.
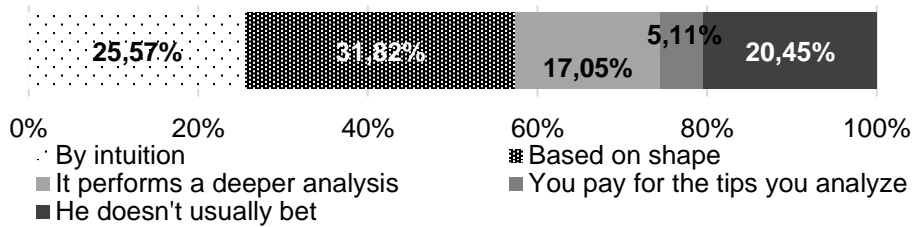
Figure 3: Decision rate of sports betting outcome
Source: Own editing

The analysis revealed whether the economic situation caused by the epidemic would increase the role of data in the selection of new players. 63% of respondents say the role of data is becoming more valuable in all teams. Buy underrated (high-potential) footballers instead of overrated players. 12% of respondents say the epidemic does not affect transfers in football at all.

In conclusion

The policies of Brentford and Midtjylland presented in our article are very good examples of the value of investing in 2-3 quality player watches instead of a foreign football player. It is advisable to look for scouts who can interpret the data well and use it to discover affordable players with high potential (whether Hungarian or foreign) who are professionally advanced in such a certification. This would benefit clubs and players alike. For players, this is good because undervalued players would be discovered and given the opportunity for continuous improvement. This is good for clubs because they could make a profit from reselling players and later operate on a market basis. With this, the club can stabilize its financial background and certify additional quality players.

Based on the questionnaire survey, it can be concluded that the majority of football professionals, footballers and fans are aware of the importance of statistical indicators. Respondents are familiar with the companies that deal with these. The majority of them know (and use) the newly released statistical indicators in addition to the traditional statistical indicators. The majority of professionals, footballers and fans are happy to use statistics for information, evaluation and analysis (even for sports betting purposes). It is clear from the research that professionals, footballers and supporters agree that due to the economic situation caused by the epidemic, club teams need to rely much more on data-driven player monitoring. Based on these, you need to invest wisely in players (you shouldn't contract overvalued footballers for huge money). Respondents also agreed that it is not possible to predict the outcome of a match with traditional statistical indicators (goal shots, corners, possession of the ball, etc.). These statistics do not reflect the exact position of a match.

References
1. I01. WORLD LEADERS IN SPORT DATA. Detailed sports data from Stats Perform to bring your content and coverage to life. https://www.statsperform.com/opta/ [06 Jan 2022]
2. I02. Unibet Magyarország. https://www.youtube.com/channel/UCFT5nbb-hRrINtwmQaWRogg [12 Dec 2021]
3. I03. Futballszótár. https://nagykero.wordpress.com/futball-szotar/ [03 Jan 2022]
4. bunteto.com
5. EU (2021). Why Bother? Higher Values and the Universal Cause in Teaching the EU. ISBN 978-80043-275-8
6. transfermarkt.de. Europa League. https://www.transfermarkt.de/ [20 Feb 2022]
7. Football statistics (2022). www.21stclub.com [19 Feb 2022]
8. www.nemzetisport.hu