

STATISTICKÉ POROVNÁNÍ ČESKÉ A DÁNSKÉ FOTBALOVÉ LIGY: VĚKOVÉ A NÁRODNOSTNÍ SLOŽENÍ HRÁČŮ

STATISTICAL COMPARISON OF CZECH AND DANISH FOOTBALL LEAGUES: MARKET VALUE, AGE, AND NATIONALITY OF PLAYERS

M. Tomíček, & N. Pelloneová

Technická univerzita v Liberci, Ekonomická fakulta, Katedra podnikové ekonomiky a managementu

Abstract

Football is one of the most popular sports disciplines and the use of statistical methods to analyze sports data is of great interest not only to coaches but also to researchers. Football clubs mainly use the services of specialized companies for data analysis. With the use of statistical analysis, it is possible to find out interesting information and relationships that are not possible by simply observing the game or the players. There is a great interest in sports data analysis among football clubs. The aim of this paper is to use statistical analysis to compare two professional European football competitions and to identify their main similarities and differences in three highlighted areas - age of players, national composition of competitions and market value of players. Statistical data analysis methods were used to test the research questions formulated in this research. The database used was created based on variables obtained from InStat and Transfermarkt. The statistical program IBM SPSS Statistics 28 was used for the calculations.

Keywords: football; sports statistical analysis; market value; Fortuna:Liga; 3F Superliga

Souhrn

Fotbal je jednou z nejpobulárnějších sportovních disciplín a využití statistických metod k analýze sportovních údajů je předmětem velkého zájmu nejen trenérů ale i vědeckých pracovníků. Fotbalové kluby využívají k datové analýze především služeb specializovaných společností. S využitím statistické analýzy je možné zjišťovat zajímavé informace a souvislosti, které není možné zjistit pouhým pozorováním hry nebo hráčů. O sportovní datové analýzy je mezi fotbalovými kluby veliký zájem. Cílem tohoto příspěvku je s pomocí statistických analýz porovnat dvě profesionální evropské fotbalové soutěže a určit jejich hlavní podobnosti a rozdíly ve třech vytyčených oblastech – věk hráčů, národnostní složení soutěží a tržní hodnota hráčů. K ověření výzkumných otázek formulovaných v tomto výzkumu byly použity metody statistické analýzy dat. Databáze využitých dat byla vytvořena na základě proměnných získaných od společností InStat a Transfermarkt. K výpočtům byl použit statistický program IBM SPSS Statistics 28.

Klíčová slova: fotbal; sportovní statistická analýza; tržní hodnota; Fortuna:Liga; 3F Superliga

Introduction

The use of statistical methods is becoming an increasingly important topic of research in sports. The present research focuses on the use of statistical methods to compare football clubs in two selected professional European competitions. Football is a complex and dynamic team sport where team performance combines the interaction of technical, tactical, physical, and mental factors (Yi et al., 2019). Almost all professional football clubs use a wide range of data in their decision-making processes. The extent and manner of working with data can vary significantly from one football club to another. Football clubs do not make their exact practices of using sports data and statistics publicly known.

Large football clubs tend to go down the route of in-house data specialists, while small clubs opt for a form of conventional data software and video analysis. Nowadays, football clubs are working with many platforms to track players' sports performance and other analytical tools to reduce the potential risk of failure. It is also now possible to use algorithms that automatically detect a team's playing patterns (Harper, 2021). Some of the leading football club and player data analytics companies include InStat, Wyscout, and StatsBomb (Tomíček, 2021). Most articles dealing with statistical analysis of sports apply more than two statistical methods, and the most used statistical methods include the T-test (Calleja-González et al., 2015), ANOVA (Gardasevic & Bjelica, 2020; Konter & Yurdabakan, 2010; Konter, 2010; Lago-Ballesteros & Lago-Peñas, 2010), correlation analysis (Parker, 2016; Redkva et al., 2018; Sandgren et al., 2013; Swanepoel & Swanepoel, 2016; Yue et al., 2014), regression analysis (Oberstone, 2009; Tomíček 2021; Tomíček & Pelloneová, 2022) and cluster analysis (Moura et al., 2014; Tomíček 2021; Pelloneová & Tomíček, 2022).

The presented research focuses on a comparison of the Czech Fortuna:Liga and the Danish 3F Superliga. For the selected competitions, there is an assumption of similarity in the playing styles of the football clubs that participated in the sports competitions in the period under study. The present research extends previous research by Tomíček and Pelloneová (2022) and Tomíček (2021), which focused mainly on the statistical comparison of offensive and defensive characteristics of the above-mentioned football competitions. From the comparison made, the Danish 3F Superliga came out as less competitive, more creative and more attractive for the spectator in terms of more goals scored. For both football leagues, there was also a positive moderate dependence of points gained on the total number of fans in the stadium. Research by Tomíček and Pelloneová (2022) concluded that Czech clubs grew their total points at the end of the season with the growth of total attendance faster than Danish football clubs. The aim of this research is to extend the previous statistical examination of the similarities and differences between the Czech and Danish leagues with additional parameters. The present research focuses mainly on the players themselves and tries to identify the main differences in the market value of the players and in the age and nationality composition of the above-mentioned football competitions.

Research on the link between market value of players and player skills has been conducted by Kirschstein and Liebscher (2018), Kim et al. (2019), Oberstone (2009) and others. Kirschstein and Liebscher (2018) use robust statistical methods to investigate the dependence of a player's market value on his football skills. Oberstone (2009) uses a linear multiple regression model in his research to define six statistically significant factors that are key to a club's ultimate success as measured by points gained in the league table. Felipe et al. (2020) analyse different variables (age, position or team and league quality) and economic evaluation of players in the top European leagues. Transfermarkt crowdsourcing values were used for the economic valuation.

Many authors also deal with the question of the composition of teams in terms of nationality. The proportion of foreign players in professional football teams has been steadily increasing over the last thirty years. The cultural diversity of professional football teams has increased significantly, especially in Europe (Poli, 2016). Recruiting players regardless of their nationality allows football clubs to tap into a global talent pool and combine the specific strengths of individuals with different cultural backgrounds. At the same time, however, it presents football clubs with the challenge of players who speak different languages and have ingrained different football philosophies. For example, Penn (2016) examined football clubs in England and Italy between 1930 and 2010 in three dimensions - the origins of players, managers (coaches) and owners. Littlewood et al. (2011) examine trends in player acquisition in five major European football leagues between the 2004/05 and 2008/09 seasons. These were the leagues (English Premier League, Ligue 1, Bundesliga, Serie A and La Liga). The study examines the presence of domestic players and foreigners within the above leagues. The results suggest an increasing rate of footballer migration. Players come mainly from South America and Africa. Although the number of domestic players declined slightly, four of the five leagues remained dominated by domestic players. The German Bundesliga was the only league that had more foreign players than domestic players. Maderer et al. (2014) examine the influence of various cultural factors on team success in the five largest European football leagues (England, France, Germany, Italy, and Spain). Their research found a negative effect of team cultural diversity and coach intercultural experience on team performance.

Some authors also examine the age structure of football teams. For example, Metelski (2021) tries to identify the main factors that differentiate the value of football players. The results show that most transfers take place mainly among players aged 21-24. In addition, research has observed the highest transfer amounts in the group of players aged 21 and younger. Many clubs engage young players primarily because a young player is easier to train and adapt to the club's system.

Methods and data

The aim of the present research is to compare two selected European football competitions and to identify the main similarities and differences between them. The research seeks to identify the main differences in the market value of players and in the age and nationality composition of both football competitions. In the article, mainly statistical methods are used, which include basic methods of descriptive statistics and methods of dependence analysis. Each of the methods used serves to understand the data from a different perspective. Descriptive statistics describe and approximate the characteristics of the data under study, while dependency analyses make it possible to answer both simple and complex questions and to generalize conclusions at a given level of significance. In the present research, the four research questions below will be examined.

Which of the examined football competitions has players with higher market value?
 Which of the examined football competitions has a larger number of foreign players?
 Which of the examined football competitions has a larger number of young players?
 Does the market value of players depend on the value of the InStat index?

The above questions will be answered in the research results using the statistical methods described below.

Research set and data source

Two European football competitions with potential similarities were selected for the research - the Czech Fortuna:Liga and the Danish 3F Superliga. Both competitions are among the top professional football competitions in their respective countries. The statistical data used in this research comes from the official databases of the two football competitions under study and are supplemented by private databases of companies from the football environment. Data from InStat (InStat, 2020), a company that analyses the performance of athletes and sports teams, was used for the research. InStat Ltd. works with more than 1,600 professional sports clubs. It provides match data and detailed statistics on individual players. The aggregated data is the so-called InStat index, which summarizes player performance into a single figure. This allows easy comparison of players (InStat, 2020). Furthermore, the databases of Transfermarkt.com (Transfermarkt, 2021), the Czech Fortuna:Liga (Fortuna:Liga, 2021) and the Danish 3F Superliga (3F Superliga, 2021) were used. The Transfermarkt.com database was mainly used to assess the market value of players in each season (Transfermarkt, 2021).

The research period covers a total of five seasons. The research period was limited by InStat data for the Danish top competition, where the oldest possible sports data reported is for the 2015/16 season. For the Czech top competition it is possible to obtain data from the 2013/14 season onwards, but for the purpose of comparing the two sports competitions this data is not used in the research. The research period does not include more recent seasons because the 2020/21 season was significantly affected by the Covid-19 pandemic and some data could be distorted by this pandemic. Another reason for this is the incompleteness of more recent data for the Czech top football competition at the time of writing.

The similarity of these competitions is based on the identical game model and also on the similar position in the UEFA league coefficients (UEFA, 2021). Denmark was two places ahead of the Czech Republic (i.e. 13th) in the UEFA league coefficient after the 2019/20 season. The similar ranking in the UEFA league coefficient table refers to the similar quality of these football competitions. The UEFA league coefficients for the examined competitions are shown in Table 1.

The present research focused only on the Czech and Danish football leagues, where the financial structure and the level of the individual teams in the two selected competitions are at a comparable level. Comparing the Czech Fortuna:Liga with the best European football leagues, whose teams have

several times higher budgets and better and higher quality sports facilities, would require a different way of evaluation.

Tabulka 1./ Table 1.

Koeficienty ligových soutěží členských asociací UEFA./ League coefficients of UEFA member associations.

#	League competition	Coefficients in individual seasons					Total
		2015/16	2016/17	2017/18	2018/19	2019/20	
1.	Spain	23.928	20.142	19.714	19.571	18.928	102.283
2.	England	14.250	14.928	20.071	22.642	18.571	90.462
3.	Germany	16.428	14.571	9.8570	15.214	18.714	74.784
4.	Italy	11.500	14.250	17.333	12.642	14.928	70.653
5.	France	11.083	14.416	11.500	10.583	11.666	59.248
...	...	...	...	...	...	...	...
12.	Austria	3.800	7.375	9.750	6.200	5.800	32.925
13.	Denmark	5.500	8.500	5.250	4.875	5.125	29.250
14.	Scotland	3.000	4.375	4.000	6.750	9.750	27.875
15.	Czech Republic	7.300	5.500	5.500	6.500	2.500	27.300
16.	Cyprus	3.000	5.500	7.000	6.125	5.125	26.750

The number of teams in the Czech Fortuna:Liga did not change in the period under review. Since the 2018/19 season, a superstructure has been added to the game model. As shown in Table 2, this increased the number of games played to between 32 and 35 for teams playing in the UEFA Champions League play-offs and just 35 for teams playing in the title and relegation groups.

Tabulka 2./ Table 2.

Porovnání účastníků české a dánské fotbalové ligy./ Comparison of Czech and Danish football leagues.

Indicator	Fortuna:Liga		3F Superliga	
	until 2017/18	from 2018/19	until 2015/16	from 2016/17
No. of participants in the season	16	16	12	14
Matches in the season	30	32-35/35/35	33	27-31/32/36
of which the basic part	30	30	0	26
of which title group	0	5	0	10
of which group for Europe	0	2 to 5	0	1 to 5
of which group for rescue	0	5	0	6
No. of clubs in the period under review	20		18	
of which	Bohemia	13	Midtjylland	6
of which	Moravia	5	Hovedstaden	5
of which	Silesia	2	Syddanmark	4
of which	-		Nordjylland	3
of which	-		Sjælland	0

The first season (2015/16) of the 3F Superliga had only 12 teams and it was played without superstructure in a three-match system. Each team played 33 games in the season. From the following season (2016/17), the number of participants increased to 14 and a superstructure was added to the Danish league's game model, two seasons before Fortuna:Liga took a similar step. This move did not lead to an across-the-board increase in the number of games per season for all teams, only for teams playing in the title group. For teams participating in the relegation group or the group for the Championship (played similarly to the Fortuna:Liga play-off system), the number of matches per season has paradoxically been reduced to 27-31 and 33 matches respectively.

In the period under review, 20 teams from Bohemia, Moravia, and Silesia participated in the Czech Fortuna:Liga. Their regional distribution can be seen in the lower part of Table 2. It shows that 65 % of all teams come from Bohemia, 25 % from Moravia, and only 10 % from Silesia. This distribution roughly corresponds to the ratio of the area of these three historical parts of the Czech Republic to its total area: 66% Bohemia, 28.5% Moravia, and 5.5% Silesia. Silesia, however, has almost twice as many teams in the top competition as this part of the Czech Republic should have according to the area ratio. It is thus the second strongest "macroregion" of the Czech Republic after the capital city. The Kingdom of Denmark is administratively divided into 5 regions. Except for Sjælland, they are all represented in the top football competition by at least one team. A total of 6 of the 18 (33.33%) teams in the Danish league came from the Midtjylland region, which corresponds to the region's area, which covers over 30% of the total area of the Kingdom of Denmark. The second highest number of teams, a total of 5 (27.78 %), came from the capital region, the so-called Hovedstaden region. This is the smallest region in Denmark, occupying only 5.98 % of the Danish territory, but at the same time, Hovedstaden is the most populous region in Denmark. Approximately 32 % of the population lives there. Interestingly, Sjælland does not have a representative in the top Danish football league, but according to the ratio of population and area, there should be two to three clubs from this region in the top Danish competition.

Statistical methods

For a basic overview of the statistical data set, mainly descriptive statistical methods are used. In the research, a large volume of data from selected football competitions was worked with. Players are evaluated from a variety of statistical metrics (e.g. mean, median, variance range). Correlation and regression analysis were used in the dependency analysis. Correlation indicates the degree of association between two variables. It is also very important to distinguish between correlation and causation, i.e. a direct relationship between cause and effect. According to Hendl (2012), correlation does not imply proof of a causal relationship, i.e. that changes in one variable affect changes in the other variable. In this paper, Pearson's correlation coefficient r was mainly used. The Pearson correlation coefficient was chosen because the Shapiro-Wilk test showed that all of the variables studied had normal distributions. The coefficient r can be determined using the covariance s_{xy} and the standard deviations s_x and s_y of the two variables x and y , see relationships (1) and (2) below (Brase et al., 2019).

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad (1)$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (2)$$

Relation (1) states that the correlation coefficient has a positive value when the association of variables is positive. Conversely, when the association of variables is negative, the correlation coefficient has a negative value. The correlation coefficient r captured by relation (2) takes values $<-1;1>$. If the value of r is close to zero, then the variables x and y are uncorrelated (Hendl, 2012).

The second method is regression analysis. According to Hindls (2007), regression analysis is concerned with examining and evaluating the association between two or more statistical features. The analysis aims to penetrate the essence of the observed phenomena in a given area and to get closer to the causal relationships. The research uses the exponential regression model. Exponential regression represents the most well-known and commonly used regression function that is nonlinear in parameters (Anděl, 2007). The least squares method is not directly used to estimate the parameters of nonlinear regression functions. It can be used at the beginning of the estimation and then refine these estimates according to the methods that are usually included in statistical programs (Hindls, 2007). All calculations performed in the research were done using IBM SPSS Statistics 28 (IBM, 2023).

Research results

Among the indicators examined within the framework of basic statistical characteristics were primarily the InStat index and the market value of players (reported in thousands of euros). For all

players, their age and the number of games played per season were also determined. The aforementioned characteristics for the Fortuna:Liga and 3F Superliga players divided into field players and goalkeepers are shown in the following Table 3.

Tabulka 3./ Table 3.

Základní statistické charakteristiky vybraných fotbalových soutěží./ Basic statistical characteristics of selected football competitions.

Indicator	Metrics	Fortuna:Liga		3F Superliga	
		Players in the field	Goalkeepers	Players in the field	Goalkeepers
InStat index	Total	427 565	42 187	383 578	32 139
	Average	215.40	234.37	222.11	231.22
	Modus	209	228	225	227
	Median	214	236	224	231
	Var. range	187	119	235	101
Market value (thous. €)	Total	1 046 584	82 675	1 020 650	72 950
	Average	527.25	459.31	590.99	524.82
	Modus	200	200	400	200
	Median	300	300	350	300
	Var. range	11 975	5 900	12 950	4 950
Age (years)	Average	26.14	29.67	24.69	26.75
	Modus	26	28	24	27
	Median	26	29	24	26
	Var. range	26	23	22	18
Matches played	Average	16.44	14.09	17.49	16.99
	Modus	1	1	1	1
	Median	17	12	18	14
	Var. range	35	34	37	36

Due to the higher absolute number of clubs and thus players in the Czech top football competition, both field players and goalkeepers in the Czech Fortuna:Liga have a higher InStat index value. The average value and modus of the InStat index variable are higher for field players and goalkeepers in the Danish 3F Superliga, which are quantitatively fewer compared to the Czech Fortuna:Liga, but achieve on average better sports performance values according to the InStat index metric. However, there is a higher margin of variation for players in the field of the Danish competition than for players in the field of the Czech competition. Therefore, there is a greater difference between the best and the worst-ranked player in terms of the InStat index value.

The absolute values of the market value indicator seem to indicate that players and goalies in the Czech league are more valuable compared to the Danish league. However, the Czech Fortuna:Liga has 16 teams, whereas the Danish 3F Superliga offers a comparison of only 12 teams and, as of the 2016/17 season, 14 teams in total. This affects the total number of players that are in the selected competitions over the period. On average, the players in the Danish 3F Superliga are worth more than 60,000 euros per player. The modus and median come out higher for field players in the case of the Danish 3F Superliga.

The mean, modus, and median of age show that players and goalkeepers in the Czech Fortuna:Liga are older compared to the Danish 3F Superliga. The Danish 3F Superliga prefers younger players, which is confirmed by the smaller variance range.

The values of the arithmetic mean and median of the variable matches played show that field players and goalkeepers played more matches in the Danish top football league than players and goalkeepers in the Czech top football league. The data also show a lower rate of variation for field players (35) and goalkeepers (34) in the Danish 3F Superliga compared to field players (37) and goalkeepers (36) playing in the Czech Fortuna:Liga.

The data captured in Table 3 showed a similar market value of players in both competitions. What is different is the national composition of the most valuable players over the examined period. The top five positions for the most valuable players in the Czech league are occupied only by players from the Czech Republic and Slovakia. On the other hand, the Danish league offers two players from Denmark

in the top five most valuable players, as well as one each from Norway, Brazil, and Ghana. Despite the lower number of teams (and thus the lower number of players), the Danish league has a higher proportion of foreigners among field players (36.19% vs. 28.06%) and goalkeepers (47.48% vs. 16.67%) than the Czech league. The absolute values are shown in Table 4. The Danish league offered younger (under 25) field players and goalkeepers in the five seasons under study (despite the lower total number of field players and goalkeepers). On average, 190 players under the age of 25 played in the Danish 3F Superliga. In contrast, the Czech Fortuna:Liga offered an average of 162 players under the age of 25 per season.

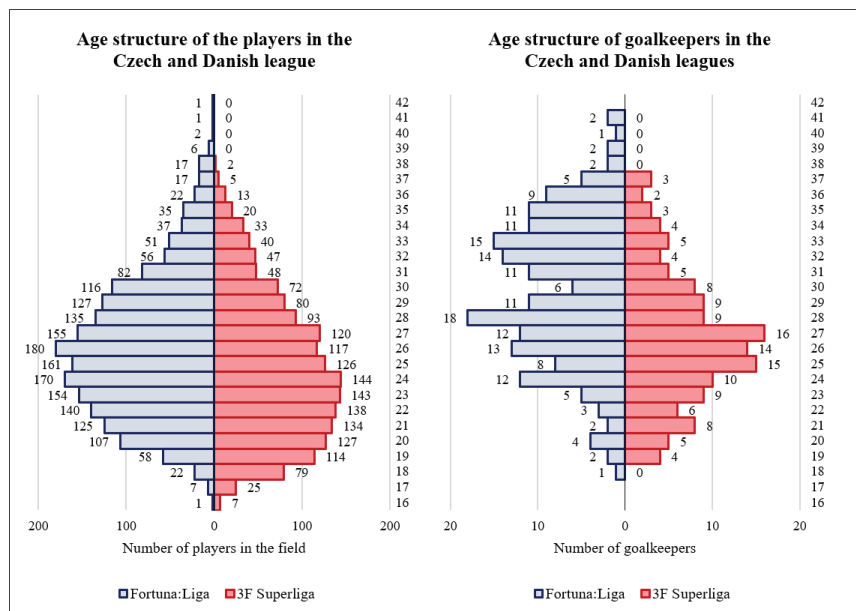
Tabulka 4./ Table 4.

Porovnání účastníků české a dánské fotbalové ligy – hráči a brankáři./ Comparison of Czech and Danish football league participants - players and goalkeepers.

Indicator	Fortuna:Liga		3F Superliga	
	Abs. values	%	Abs. values	%
Field players	1 985		1 727	
of which domestic players	1 428	71.94	1 102	63.81
of which foreigners	557	28.06	625	36.19
Goalkeepers	180		139	
of which domestic goalkeepers	150	83.33	73	52.52
of which foreigners	30	16.67	66	47.48
Field players	1 985		1 727	
of which < 25 years	784	39.50	911	52.75
of which ≥ 25 years	1 201	60.50	816	47.25
Goalkeepers	180		139	
of which < 25 years	29	16.11	42	30.22
of which ≥ 25 years	151	83.89	97	69.78

Obrázek 1./ Figure 1.

Věková struktura – hráči v poli a brankáři./ Age structure - field players and goalkeepers.



An interesting insight is provided by the following diagrams (see Figure 1), which show the age structure of players in both football competitions in more detail. The diagrams confirm the conclusions from the overview in Tables 3 and 4. Players in the Czech Fortuna:Liga were on average older in both

the field players and goalkeepers in the period under review. The Danish 3F Superliga teams preferred younger players on average, which is confirmed by the smaller variance (see Table 3). In the case of field players the maximum age of a player was 38 years and in the case of goalkeepers 37 years. In the Czech Fortuna:Liga there were field players and goalkeepers over the age of 40. One of the factors for the younger composition of the Danish league teams compared to the Czech league may be the different phase of "generational renewal" of each team. Other reasons may include the general trend in the Danish league to favor younger players over older, more experienced ones.

Correlation and regression analyses were performed separately for field players and goalkeepers, due to the different determination of the InStat index value for goalkeepers compared to other players. The results of the correlation analysis of the dependence of the market value on the InStat index value for players in the Czech Fortuna:Liga and Danish 3F Superliga in each season are summarized and presented in the following Table 5. For each season and each competition, information about the absolute frequency of players, Pearson correlation coefficient, R-squared value (coefficient of determination) determining the quality of the selected regression analysis model was added. The last column shows the prescription of the regression line equation.

Tabulka 5./ Table 5.

Korelační koeficienty pro hráče v poli v jednotlivých sezónách./ Correlation coefficients for players in the field in each season.

Group x_i	Absolute frequency n_i	Pearson correlation coefficient	Model fit R-squared R^2	Equation of the regression line
F:L players 2015/16	379	0.760449	57.83 %	$y_{1CZ} = 0.7294e^{0.0276x}$
3FS players 2015/16	308	0.509558	25.96 %	$y_{1DK} = 5.6541e^{0.0187x}$
F:L players 2016/17	404	0.338711	11.47 %	$y_{2CZ} = 18.635e^{0.0135x}$
3FS players 2016/17	343	0.468447	21.94 %	$y_{2DK} = 7.0129e^{0.0177x}$
F:L players 2017/18	380	0.474901	22.55 %	$y_{3CZ} = 1.4296e^{0.0257x}$
3FS players 2017/18	348	0.508628	25.87 %	$y_{3DK} = 1.2872e^{0.0252x}$
F:L players 2018/19	399	0.551226	30.39 %	$y_{4CZ} = 0.4165e^{0.0314x}$
3FS players 2018/19	360	0.447069	19.99 %	$y_{4DK} = 1.7511e^{0.0245x}$
F:L players 2019/20	423	0.889272	79.08 %	$y_{5CZ} = 0.5522e^{0.0296x}$
3FS players 2019/20	368	0.617616	38.14 %	$y_{5DK} = 2.2821e^{0.0235x}$
Total	3 712	x	x	x

Tabulka 6./ Table 6.

Korelační koeficienty pro brankáře v jednotlivých sezónách./ Correlation coefficients for goalkeepers in each season.

Group x_i	Absolute frequency n_i	Pearson correlation coefficient	Model fit R-squared R^2	Equation of the regression line
F:L goalkeepers 2015/16	35	0.548081	30.04 %	$y_{1CZ} = e^{0.0164x}$
3FS goalkeepers 2015/16	25	0.549170	30.16 %	$y_{1DK} = e^{0.029x}$
F:L goalkeepers 2016/17	33	0.686449	47.12 %	$y_{2CZ} = e^{0.0313x}$
3FS goalkeepers 2016/17	29	0.404959	16.40 %	$y_{2DK} = e^{0.0136x}$
F:L goalkeepers 2017/18	39	0.571061	62.61 %	$y_{3CZ} = e^{0.0285x}$
3FS goalkeepers 2017/18	28	0.504197	25.42 %	$y_{3DK} = e^{0.0441x}$
F:L goalkeepers 2018/19	34	0.364963	13.32 %	$y_{4CZ} = e^{0.0162x}$
3FS goalkeepers 2018/19	27	0.726283	52.75 %	$y_{4DK} = 0.0058e^{0.0476x}$
F:L goalkeepers 2019/20	39	0.520197	27.06 %	$y_{5CZ} = e^{0.0248x}$
3FS goalkeepers 2019/20	30	0.770163	59.32 %	$y_{5DK} = e^{0.032x}$
Total	319	x	x	x

The above data show that when the files are split by seasons, the quality of the regression model is low in most cases. In only two cases does it exceed the 50 % threshold. However, the Pearson correlation coefficient is around 0.50 in all the seasons examined in both competitions, indicating a moderately strong relationship.

The same principle was applied to the correlation and regression analyses in each season for goalkeepers. Table 6 captures information on the absolute frequency of the group studied, the value of the Pearson correlation coefficient, and the quality level of the selected regression model. The last column contains the equation prescription of the regression line.

Due to the very low number of observations (25 to 39) in the analysis of goalies in each season, the parameter b_0 is not significant. The exception is the analysis of Danish league goalkeepers in the 2018/19 season. The Pearson correlation coefficient ranges from 0.36 to 0.77. This is therefore a moderate to very strong correlation. The quality of the regression model (R-squared) is similar to the regression analysis of players in the field in individual seasons. The highest value of the coefficient of determination was achieved in the analysis of goalkeepers in the Czech Fortuna:Liga in the 2017/18 season. On the contrary, the lowest value of the coefficient of determination can be observed in the 2018/19 season of the goalkeepers of the Danish 3F Superliga.

Discussion and conclusion

The presented research aimed to compare two European football competitions using statistical indicators. The research mainly used the methods of descriptive statistics, correlation, and regression analysis. The analyzed football competitions Fortuna:Liga and 3F Superliga were selected based on of similar playing models, financial frameworks of the participating clubs, and their playing styles.

The initial statistical characterization of the selected football competitions began, among other things, by comparing the game metrics of the two competitions. The Danish 3F Superliga had more foreign players (also in absolute terms) and young players under 25 years of age. The results of the research in the case of the Danish 3F Superliga are consistent with the findings of Metelski (2021), who states that football clubs engage young players because of the better opportunity to train players and adapt to the club's system. For the Danish 3F Superliga, Poli's (2016) assertion that, especially in Europe, the cultural diversity of professional football teams is increasing has also been confirmed. The Danish 3F Superliga players were on average more valuable in terms of market value compared to the Czech Fortuna:Liga players, as shown by the higher interval modus than the Czech players. This research therefore found answers to the first three research questions, which related to market value, national composition and age of players.

Correlation and regression analyses were performed to answer the final research question, whether and how the market value of a player depends on his sports performance (as measured by the InStat index). Multiple periods were considered for both competitions separately. A regression model with an exponential curve was chosen based on the data. It was compared with the linear curve, which did not reach the corresponding quality values of the exponential curve in any of the tests performed. Using an exponential curve means that a player's value increases proportionally faster than his game performance as measured by the InStat index. The exponential model corresponds to the trend of recent years in the football environment. It is an increase in the market values of the so-called difference players, who can decide matches with their performances. For these players, however, the increase in market value is non-linear. The increase in the market value of the best players "pulls" the market values of the other players upwards. The research reaches a similar conclusion to Kirschstein and Liebscher (2018), suggesting that a player's market value depends on his football skills and that star players tend to be overvalued.

In the first observed season 2015/16, we can observe a different comparison between the two competitions for field players and goalkeepers in terms of the growth rate of market value depending on the value of the InStat index. The 2015/16 season is almost identical for goalkeepers in both competitions in terms of R^2 value determining the quality of the regression model, which is around 30%. The set of field players in the Czech Fortuna:Liga (57.83%) showed twice the R^2 value. The 2016/17 season shows the lowest correlation coefficient (0.338711) and coefficient of determination (11.47%) for players in the Czech Fortuna:Liga. This is the "least quality" use of the model of all the observations. In the case of goalkeepers in both competitions, a noticeable difference in terms of the coefficient of determination can be observed in the 2016/17 season, while the data for Czech league goalkeepers is described by the model as 47%, and for Danish goalkeepers, it is only 16%. The coefficient of determination in the 2017/18 season for the regression analysis of field players is 23% (Fortuna:Liga) and 26% (3F Superliga). This is a slight increase from the previous 2016/17 season. Regression analysis

of the 2018/19 season for Czech Fortuna:Liga players are accompanied by an increase in the coefficient of determination to 30% and a decrease to less than 20% for Danish 3F Superliga players. Regression analysis of Czech Fortuna:Liga goalkeepers revealed the lowest coefficient of determination (13.32%) this season. Goalkeepers in the Czech Fortuna:Liga have the lowest correlation coefficient (0.364963) in the 2018/19 season of all seasons examined. On the other hand, the correlation coefficient of Danish 3F Superliga goalkeepers is above 0.7, indicating a strong correlation between market value and the InStat index. For the last season 2019/20, a higher value of both the correlation coefficient and the coefficient of determination is typical for all playing positions. One of the factors behind the relatively high values of the correlation coefficient and the coefficient of determination is the average increase in the market value of players in both competitions. This is reflected in the increased quality of the exponential regression model.

In conclusion, there were more foreign and young players under 25 in the Danish 3F Superliga. The Danish 3F Superliga players were on average more valuable in terms of market value than the Czech Fortuna:Liga players. It was also found that a player's market value grows proportionally faster than his game performance as measured by the InStat index.

References

- 3F Superliga. (2021). *Superliga.dk: Performance Center*. <https://www.superliga.dk/performance-center-superliga-2020-2021>
- Anděl, J. (2007). *Statistické metody*. Matfyzpress.
- Brase, Ch. H., Brase, C. P., & Iraki, A. (2019). *Understanding basics statistics*. Cengage.
- Calleja-González, J., Los Arcos, A., Mejuto, G., Casamichana, D., San Roman Quintana, J., & Yanci, J. (2015). Reproducibilidad de test de aceleración y cambio de dirección en fútbol. *Revista Internacional de Ciencias del Deporte*, 11(40), 104–115. <http://dx.doi.org/10.5232/ricyde2015.04001>
- Felipe, J. L., Fernandez-Luna, A., Burillo, P., de la Riva, L. E., Sanchez-Sanchez, J., & Garcia-Unanue, J. (2020). Money talks: Team variables and player positions that most influence the market value of professional male footballers in Europe. *Sustainability*, 12(9), 3709. <https://doi.org/10.3390/su12093709>
- Fortuna:Liga. (2021). *FortunaLiga.cz: Statistiky*. <https://www.fortunaliga.cz/statistiky?unit=1¶meter=1>
- Gardasevic, J., & Bjelica, D. (2020). Body Composition Differences between Football Players of the Three Top Football Clubs. *International Journal of Morphology*, 38(1), 153–158. <https://doi.org/10.4067/S0717-95022020000100153>
- Harper, J. (2021). *Data experts are becoming football's best signings*. BBC News. <https://www.bbc.com/news/business-56164159>
- Hendl, J. (2012). *Přehled statistických metod: analýza a metaanalýza dat*. Portál.
- Hindls, R. (2007). *Statistika pro ekonomy*. Professional Publishing.
- IBM. (2023). *IBM SPSS software*. <https://www.ibm.com/spss>
- InStat. (2020). *Objective performance rating for players & teams*. <https://instatsport.com/football>
- Kim, Y., Nam Bui, K., & Jung, J. J. (2019). Data-driven exploratory approach on player valuation in foot-ball transfer market. *Concurrency and Computation Practice and Experience*, 33(3). <https://doi.org/10.1002/cpe.5353>
- Kirschstein, T., & Liebscher, S. (2018). Assessing the market values of soccer players - a robust analysis of data from German 1. and 2. Bundesliga. *Journal of Applied Statistics*, 46(7), 1336–1349. <https://doi.org/10.1080/02664763.2018.1540689>
- Konter, E., & Yurdabakan, I. (2010). Nonverbal intelligence of soccer players according to their age, gender and educational level. *Social and Behavioral Sciences*, 2(2), 915–921. <https://doi.org/10.1016/j.sbspro.2010.03.126>
- Konter, E. (2010). Nonverbal intelligence of soccer players according to their level of play. *Social and Behavioral Sciences*, 2(2), 1114–1120. <https://doi.org/10.1016/j.sbspro.2010.03.126>
- Lago-Ballesteros, J., & Lago-Peñas, C. (2010). Performance in Team Sports: Identifying the Keys to Success in Soccer. *Journal of Human Kinetics*, 25(2010), 85–91. <https://doi.org/10.2478/v10078-010-0035-0T>

- Littlewood, M., Mullen, C., & Richardson, D. (2011). Football labour migration: An examination of the player recruitment strategies of the 'Big Five' European Football Leagues 2004–5 to 2008–9. *Soccer & Society*, 12(6), 788–805. <https://doi.org/10.1080/14660970.2011.609680>
- Maderer, D., Holtbrügge, D., & Schuster, T. (2014). Professional Football squads as multicultural teams. *International Journal of Cross Cultural Management*, 14(2), 215–238. <https://doi.org/10.1177/1470595813510710>
- Metelski, A. (2021). *Journal of Physical Education and Sport*, 21(2), 1150–1155. <https://doi.org/10.7752/jpes.2021.s2145>
- Moura, F. A., Barreto Martins, L. E., & Cunha, S. A. (2014). Analysis of football game-related statistics using multivariate techniques. *Journal of Sports Sciences*, 32(20), 1881–1887. <https://doi.org/10.1080/02640414.2013.853130>
- Oberstone, J. (2009). Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success. *Journal of Quantitative Analysis in Sports*, 5(3). <https://doi.org/10.2202/1559-0410.1183>
- Parker, S. D. G. (2016). *Correlation Between the National Football League Draft and Player Performance*. Robert D. Clark Honors College.
- Pelloneová, N., & Tomíček, M. (2022). Ranking players by DEA: An analysis of Czech and Danish football. *Studia Sportiva*, 16(1), 76–90. <https://doi.org/10.5817/sts2022-1-8>
- Penn, R. (2018). From comparative analysis to final publication: Nationalism and Globalization in English and Italian football since 1930. *European Journal for Sport and Society*, 10(4), 345–364. <https://doi.org/10.4135/9781526445308>
- Poli, R., Ravenel, L., & Besson, R. (2016). *CIES Football Observatory Monthly Report - Foreign players in football teams*. https://football-observatory.com/IMG/pdf/mr12_eng.pdf
- Redkva, P. E., Paes, M. R., Fernandez, R., & da-Silva, S. G. (2018). Correlation between Match Performance and Field Tests in Professional Soccer Players. *Journal of Human Kinetics*, 62(1), 213–219. <https://doi.org/10.1515/hukin-2017-0171>
- Sandgren E., Karlsson, M., & Ji-Guo Y. (2013). Correlation Analysis Between SoccerGame World Ranking and Player League Distribution. *Sport and Art*, 1(2), 34–40. <https://doi.org/10.13189/saj.2013.010202>
- Swanepoel, M. J., & Swanepoel, J. (2016). The correlation between player valuation and the bargaining position of clubs in the English premier league (EPL). *International Journal of Economics and Finance Studies*, 8(1), 209–225.
- Tomíček, M., & Pelloneová, N. (2022). Sports data analysis: Comparison of the Czech and Danish top football competitions. *Tělesná Kultura*, 45. <https://doi.org/10.5507/tk.2022.002>
- Tomíček, M. (2021). *Aplikace statistických metod ve sportovních klubech* [Diplomová práce, Technická univerzita v Liberci].
- Transfermarkt. (2021). *Transfermarkt.com: Intern*. <https://www.transfermarkt.com/intern/stellenangebot>
- UEFA. (2021). *UEFA Country Coefficients. 2021, UEFA Rankings*. <https://www.uefa.com/member-associations/uefarankings/country/#/yr/2021>
- Yi, Q., Gómez, M. A., Liu, H., & Sampaio, J. (2019). Variation of match statistics and football teams' match performance in the group stage of the UEFA Champions League from 2010 to 2017. *Kinesiology*, 51(2), 170–181. <https://doi.org/10.26582/k.51.2.4>
- Yue, Z., Broich, H. & Mester, J. (2014). Statistical Analysis for the Soccer Matches of the First Bundesliga. *International Journal of Sports Science & Coaching*, 9(3), 553–560. <https://doi.org/10.1260/1747-9541.9.3.553>

Ing. Michal Tomíček
Technická univerzita v Liberci
Ekonomická fakulta
Katedra podnikové ekonomiky a managementu
Voroněžská 1329/13, 46001 Liberec
michal.tomicek@tul.cz