

# Sportovní datová analýza: Porovnání české a dánské nejvyšší fotbalové soutěže

Michal Tomíček\*, Natalie Pelloneová

*Ekonomická fakulta, Technická univerzita v Liberci, Česká republika*

Copyright: © 2022 M. Tomíček, & N. Pelloneová. Toto je open access článek vydaný pod Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

**Východiska:** Fotbal je jedním z nejpobulárnějších sportů na světě a použití statistických modelů k analýze dat je předmětem velkého zájmu trenérů, sportovních manažerů, fanoušků, novinářů a samozřejmě i samotných sportovců. V posledních letech využívají fotbalové kluby velice často služeb společností, které se zabývají analýzou sportovních dat. Pomocí datové analýzy je možné zjišťovat informace, které nelze prostým pozorováním hry zjistit. Z tohoto důvodu je v současné době o sportovní datové analýzy veliký zájem, a to především mezi fotbalovými kluby, které na základě nich činí důležitá strategická rozhodnutí. **Cíle:** Pomocí statistických analýz porovnat dvě profesionální evropské fotbalové soutěže a určit jejich hlavní podobnosti a rozdíly. **Metodika:** K ověření výzkumných otázek formulovaných v tomto výzkumu byly použity metody statistické analýzy dat. Databáze byla vytvořena na základě proměnných získaných od společnosti InStat, která se zabývá analýzou výkonnosti sportovců a sportovních týmů. K výpočtům byl použit statistický software Statgraphics Centurion XVIII. **Výsledky:** Dánský fotbal vychází ze srovnání s českým fotbalem jako méně souborový, kreativnější a pro diváka atraktivnější z pohledu většího množství vstřelených branek. Dánský fotbal má větší zastoupení cizinců mezi hráči. U obou analyzovaných soutěží byla prokázána pozitivní středně silná závislost bodového zisku na celkovém počtu fanoušků na stadionu pro klub. Českým klubům s růstem celkové návštěvnosti rostl celkový bodový zisk na konci sezony rychleji než dánským klubům. U obou analyzovaných soutěží byla prokázána středně silná korelace tržní hodnoty hráčů v poli a jejich výkonnosti. U českého hráče v poli rostla jeho tržní hodnota v závislosti na herních výkonech rychleji než u hráče dánské soutěže. **Závěry:** Ve výzkumu byly porovnány dvě evropské fotbalové soutěže s vysokým potenciálem podobnosti pomocí statistické datové analýzy. Datovou analýzu ve sportu lze doporučit v rozhodovacích procesech zejména z důvodu jejich ekonomických i sportovních benefitů.

**Klíčová slova:** fotbal; korelační analýza; regresní analýza; tržní hodnota fotbalových hráčů; sportovní analýza

## Úvod

Fotbal je komplexní a dynamický týmový sport, kde výkonnost týmů kombinuje interakce technických, taktických, fyzických a mentálních faktorů. V tomto nepředvídatelném kontextu je pro trenéry a výzkumníky velice obtížné zjistit, které faktory nejvíce přispívají k týmovému úspěchu a jak se tyto faktory vzájemně ovlivňují. Nicméně v posledních letech bylo možné díky pokroku v systémech video-analýzy získat velké množství zápasových statistik (Yi et al., 2019).

Studie zabývající se statistikou fotbalových zápasů, výkonností hráčů a týmů mají dlouhou historii. Prakticky všechny profesionální fotbalové kluby v Evropě využívají data ve svých rozhodovacích procesech,

ale míra a způsob práce s daty se mezi jednotlivými fotbalovými kluby odlišuje. Většina fotbalových klubů nedává přesné postupy využití sportovních dat a statistik veřejně najevo. Velké kluby mají sklon jít cestou interních datových specialistů, zatímco malé kluby volí formu běžného datového softwaru a analýzu videozáznamů. V současné době pracují fotbalové kluby s mnoha platformami sledování sportovní výkonnosti hráčů a dalšími analytickými nástroji, které mají za úkol snížit možné riziko neúspěchu či špatné volby. Nyní je také možné využití algoritmů, které automaticky detekují herní vzorce týmu (Harper, 2021). Mezi přední společnosti zabývající se datovou analýzou fotbalových klubů a hráčů patří např. společnosti InStat, Wyscout nebo StatsBomb.

Vhodné použití statistických metod je stále důležitější ve výzkumu v oblasti sportu. V současnosti lze mezi kluby na cestě k úspěchu spatřovat dva hlavní přístupy. V prvním případě je jednou z možných cest k úspěchu nákup hvězd do velkoklubů. Ve druhém

\*Korespondenční adresa: Ing. Michal Tomíček, Technická univerzita v Liberci, Ekonomická fakulta, Katedra podnikové ekonomiky a managementu, Studentská 1402/2, 461 17 Liberec, e-mail: [michal.tomicek@tul.cz](mailto:michal.tomicek@tul.cz)

případě je to využití sportovní datové analýzy. Malé fotbalové kluby nemohou z finančního hlediska konkurovat kapitálové síle velkoklubů a musí se tudíž spoléhat na kvalitní datovou analýzu. Předložený výzkum je zaměřen na využití druhého principu a následném použití širokého spektra statistických metod, pomocí kterých jsou analyzovány fotbalové kluby dvou vybraných profesionálních evropských soutěží.

#### *Přehled používaných statistických metod ve fotbale*

V současnosti je k dispozici celá řada recenzovaných časopisů, které se zabývají hledáním nových a inovativních metod pro analýzu sportovních dat. Každoročně je také pořádáno několik významných konferencí, které jsou věnovány prezentaci výsledků sportovního výzkumu. Stejně jako v mnoha jiných oborech i v oblasti sportu dochází díky novým technologiím k přílivu velkého množství sportovních dat, což vytváří příležitost pro pokročilé statistické analýzy. V následující části příspěvku je nastíněn přehled vybraných autorů, kteří ve svých výzkumech aplikují různé statistické a matematické metody pomocí kterých hodnotí výkonnost fotbalových klubů a hráčů a snaží se odhalit hlavní faktory jejich úspěchů. Většina autorů ve svých výzkumech využívala více než dvě statistické metody a nejpoužívanějšími přístupy byly T-test (Calleja-González et al., 2015), ANOVA (Gardasevic & Bjelica, 2020; Konter & Yurdabakan, 2010; Konter, 2010) a korelační analýza (Parker, 2016; Redkva et al., 2018; Sandgren et al, 2013; Swanepoel & Swanepoel, 2016).

Na výzkum souvislostí mezi tržní hodnotou hráčů a hráčských dovedností se zaměřili např. Kirschstein a Liebscher (2018). Autoři zkoumali s využitím robustních statistických metod, do jaké míry závisí tržní hodnota hráče na jeho fotbalových dovednostech. Ve svém výzkumu analyzovali soubor dat zahrnující 28 výkonnostních metrik a tržní hodnotu 493 hráčů z 1. a 2. Bundesligy. Autoři Kim et al. (2019) navrhovali pomocí regresní analýzy a shlukové analýzy přístup pro klasifikaci a hodnocení fotbalistů na základě jejich předvedených výkonů. Oberstone (2009) ve svém výzkumu využíval analýzu založenou na vícenásobné regresi. Lineární vícenásobný regresní model používal k definování šesti statisticky významných faktorů, které jsou klíčem ke konečnému úspěchu klubu měřeného pomocí bodů získaných v sezóně 2007/08 anglické Premier League.

Někteří autoři např. Lago-Ballesteros a Lago-Peñas (2010) ve svém výzkumu aplikovali na získaná data analýzu rozptylu. Cílem jejich studie bylo identifikovat konkrétní výkonnostní ukazatele, které odlišují špičkové španělské fotbalové kluby od ostatních. Autoři analyzovali 380 odehraných zápasů v sezóně 2008/09, přičemž se zaměřují jak na ofenzivní ukazatele (asistence, centry, zahrané ofsajdy, přestupky ve prospěch

týmu, zahrané kopy z rohů, držení míče), tak také na defenzivní ukazatele (obdržené ofsajdy, spáchané přestupky, kopy z rohů proti, žluté karty, červené karty) a v neposlední řadě na ukazatele spojené s gólovými situacemi (góly, celkový počet střel, střely na branku, přesnost střelby). Výsledky výzkumu naznačovaly, že nejlepší týmy měly vyšší průměr vstřelených branek, celkový počet střel a střel na branku než týmy ze spodu a středu tabulky. Týmy na spodních místech tabulky potřebovaly ke vstřelení branky vyšší počet střel než ostatní skupiny týmů. Týmy ze středu tabulky zase vykazovaly nižší hodnotu asistencí a držení míče než nejlepší týmy.

Někteří autoři (např. Broich et al., 2014; Yue et al., 2014) zjišťovali, které herní parametry jsou pro výsledek zápasu důležitější. Delgado-Bordonau et al. (2013) analyzovali dopad vybraných ukazatelů ofenzivní a defenzivní výkonnosti ve vztahu k úspěchu týmů na mistrovství světa ve fotbale v roce 2010. Pro analýzu shromáždili tyto herní statistiky: celkový počet střel, střely na branku, střely mimo branku, podíl střel na bránu z celkového počtu střel, podíl střel mimo bránu z celkového počtu střel, útočná efektivita (vyjádřená jako podíl vstřelených gólů k celkovému počtu střel) a obranná efektivita (vyjádřená jako podíl obdržených gólů k celkovému počtu střel na branku). Kromě toho zkoumali také vliv prvního gólu na výsledek zápasu. Výsledky studie ukázaly, že útočné proměnné související se střelami na branku a útočnou efektivitou byly na mistrovství světa ve fotbale v roce 2010 lepšími ukazateli týmového úspěchu než obranné proměnné.

Dalším výzkumem se zabýval např. Moura et al. (2014), kteří zkoumali fotbalové herní statistiky pomocí analýzy hlavních komponent a shlukové analýzy s cílem určit, zda je možné odlišit vítězné týmy od remizujících a poražených. Výsledky výzkumu ukázaly, že 70 % vítězných týmů bylo zařazeno do stejné skupiny. Na druhé straně bylo 68 % remizujících a prohrávajících týmů zařazeno do druhé skupiny.

Broich et al. (2014) pomocí statistické analýzy 153 zápasů 1. Bundesligy porovnávali různé týmové parametry mezi vítěznými a poraženými týmy. Výsledky podporují závěry dřívější analýzy (Yue et al., 2014), že pro vítězství v zápase je gólová efektivita, definovaná jako podíl počtu gólů a počtu střel, důležitější než kvantita střel. Předložená analýza potvrdila, že kvalita střel, reprezentovaná gólovou efektivitou, je pro vítězství ve fotbalovém utkání důležitější než množství střel.

Dos Santos et al. (2016) využívali statistické nástroje pro zjištění finanční výkonnosti fotbalových klubů. Analyzovali vztah mezi žebříčkem vytvořeným Brazílskou fotbalovou konfederací a ekonomicko-finančními ukazateli brazilských fotbalových týmů. Výzkumný vzorek se skládal z 36 brazilských fotbalových týmů, které patří do sérií A, B a C. Pro analýzu dat využili

metodu multikriteriálního rozhodování VIKOR, která byla aplikována spolu s Kendallovou hodnotní korelací. Výsledky ukázaly, že většina brazilských fotbalových týmů má nedostatečnou likviditu, jsou závislé na zdrojích třetích stran, a mají negativní ziskovost.

Ve výzkumu byla porovnávána česká a dánská nejvyšší fotbalová soutěž. U vybraných soutěží existoval předpoklad podobnosti herních stylů jednotlivých týmů, které se soutěží ve zkoumaném období účastnily. Ve zkoumaných sezonách došlo v obou soutěžích ke změně herního systému. Obě soutěže zaujímají blízké postavení v pořadí evropských soutěží UEFA založeném na výkonnosti týmů v evropských klubových soutěžích.

## Metodika

Cílem předloženého výzkumu je porovnání dvou evropských fotbalových soutěží pomocí statistických analýz a určení podobností a rozdílů těchto vybraných soutěží. Použité metody zahrnovaly metody popisné statistiky, korelační a regresní analýzy.

V předkládaném výzkumu bude ověřováno sedm níže uvedených výzkumných otázek.

1. Mají hráči dánské 3F Superligy na svém kontě v průměru vyšší počet vstřelených branek v porovnání s hráči české Fortuna:Ligy?
2. Česká Fortuna:Liga a dánská 3F Superliga jsou z hlediska souborovosti odlišné?
3. V české Fortuna:Lize a dánské 3F Superlige působí odlišný počet zahraničních hráčů?
4. Jsou zkoumané soutěže odlišné z hlediska průměrné divácké návštěvnosti?
5. Je bodový zisk klubu závislý na divácké návštěvnosti?
6. Je tržní hodnota hráčů v poli závislá na jejich sportovní výkonnosti?
7. Mění se závislost tržní hodnoty hráčů v poli na jejich sportovní výkonnosti v obou soutěžích odlišným tempem?

Výše uvedené otázky budou zodpovězeny s využitím níže popsaných statistických metod a analýz v hlavních výsledcích výzkumu.

### Zdroj dat

Data použitá v tomto výzkumu pochází z oficiálních databází obou zkoumaných fotbalových soutěží a jsou doplněna o soukromé databáze společností z fotbalového prostředí. Hlavním zdrojem dat jsou data získaná od společnosti InStat (InStat, 2020), která se zabývá analýzou výkonnosti sportovců a sportovních týmů. Dále bylo ve výzkumu využito databází serveru Transfermarkt.com (Transfermarkt, 2021) a databáze mi české Fortuna:Ligy (Fortuna:Liga, 2021) a dánské

3F Superligy (3F Superliga, 2021). Kombinace několika výše jmenovaných databází v sobě skrývá výhodu v podobě komplexnosti dat, ale na druhé straně také čítá několik problémů, od nepřesností až po kumulaci možných chyb v databázích. Z tohoto důvodu je jako primární zdroj dat použita oficiální databáze dané ligové soutěže. Na ní jsou následně „nabalena“ sportovní data společnosti InStat a ze serveru Transfermarkt.com jsou hráčům přidány tržní hodnoty v jednotlivých zkoumaných sezonách.

Výzkumné období zahrnuje sezony 2015/16 až 2019/20. Výzkumné období bylo omezeno daty společnosti InStat pro dánskou nejvyšší soutěž, kdy nejstarší možná uváděná sportovní data jsou za sezonu 2015/16. Pro českou nejvyšší soutěž je možné získat data počínaje sezonou 2013/14, ale pro účely porovnání obou sportovních soutěží nebudou tato data ve výzkumu využita.

### Výzkumný soubor

Z fotbalového prostředí byly do předloženého výzkumu k porovnání vybrány dvě evropské fotbalové soutěže s potenciálem podobnosti – česká Fortuna:Liga a dánská 3F Superliga. Tyto soutěže jsou nejvyššími profesionálními fotbalovými soutěžemi ve svých zemích. Podobnost těchto soutěží vychází především z herního modelu, kdy obě soutěže adaptovaly systém skupin o titul a o záchranu. V sezonách 2015/16 až 2017/18 se v české nejvyšší soutěži hrála pouze základní část soutěže (bez nadstavby a s přímým sestupem dvou nejhorších celků). Od sezony 2018/19 přibyla do modelu nadstavba zvyšující počet zápasů ze 30 na 35 pro týmy ve skupině o titul, na 35 pro týmy ve skupině o udržení (druhý a třetí nejhorší tým ještě baráž) a na 32 až 35 pro kluby ve skupině o evropské poháry (dle toho, jak daleko se proboují – systém play-off).

Oproti tomu dánská nejvyšší soutěž adaptovala nadstavbu již v sezoně 2016/17. Vzhledem k nižšímu počtu účastníků v dánské lize jsou po základní části kluby rozděleny do dvou skupin – o titul a o udržení. Kluby ve skupině o udržení jsou rozděleny do dvou skupin, kdy první dva postupují do bojů o evropské poháry, třetí ve skupině nastoupí do baráže o udržení v nejvyšší soutěži a poslední týmy ve skupinách sestupují přímo. Jedná se o divácky pravděpodobně atraktivnější systém, kdy i týmy ze spodní poloviny tabulky mají reálnou šanci se účastnit evropských pohárů. Herní modely tedy nejsou totožné, ale v obou soutěžích je přítomný nadstavbový model.

Další podobností vybraných evropských soutěží je jejich umístění v pořadí ligových koeficientů UEFA. Blízké pořadí v tabulce ligových koeficientů UEFA značí podobné kvality nejlepších celků každé soutěže. Tyto a další podobnosti jsou předpokladem provedení statistického porovnání.

České nejvyšší fotbalové soutěže se účastní celkem 16 týmů z Čech, Moravy a Slezska. Počínaje sezonou 2018/19 nese česká nejvyšší fotbalová soutěž oficiální název Fortuna:Liga. Jedná se již pátou změnu názvu od sezony 1997/98, kdy nesla česká nejvyšší fotbalová liga poprvé název svého hlavního sponzora. Dánská 3F Superliga je nejvyšší profesionální fotbalovou soutěží v Dánsku, které se účastní celkem 12 týmů z 5 dánských regionů. Dánsko bylo po sezoně 2019/20 v ligovém koeficientu UEFA o dvě příčky před Českou republikou (tj. na 13. místě).

Předkládaný výzkum se kromě samotných fotbalových soutěží zaměřuje také na výkonnost a tržní hodnotu hráčů fotbalových klubů účastnících se výše popsaných soutěží a omezuje se pouze na hráče, kteří nastoupili alespoň k jednomu zápasu v sezoně. U hráčů, kteří nastoupili v jedné sezoně za více klubů, bylo určujícím faktorem pro přiřazení klubu počet odehraných zápasů v sezoně za dané kluby. Hráči byl následně přiřazen klub, za který odehrál v dané sezoně více zápasů, v případě rovnosti rozhodoval počet odehraných minut. V případě hostujících hráčů v rámci stejné ligové soutěže není určující „mateřský“ klub, ale klub, za který hráč v sezoně nastupoval. Hráč, který nenastoupil v sezoně k žádnému zápasu (odehrál právě nula minut), nebyl ve výpočtech a analýzách zohledněn.

Porovnání české nejvyšší fotbalové soutěže Fortuna:Ligy s nejlepšími evropskými fotbalovými ligami, jejichž týmy disponují několikanásobně vyššími rozpočty a také lepším a kvalitnějším sportovním zázemím by vyžadovalo jiný pohled na porovnávání hodnot hráčů. I proto se předkládaný výzkum zaměřil pouze na českou a dánskou fotbalovou ligu, kde je finanční struktura a úroveň jednotlivých týmů obou vybraných soutěží na srovnatelné úrovni.

#### *Statistické metody*

Pro základní přehled o souboru statistických dat jsou použity vybrané metody popisné statistiky. Ve výzkumu jsme pracovali s velkým objemem dat vybraných fotbalových soutěží. Výkony hráčů jsme sledovali z širokého spektra sportovních metrik (např. průměr, medián, variační rozpětí).

V oblasti analýzy závislostí bylo využito zejména korelačních a regresních analýz. Pomocí těchto analýz lze poskytnout číselné charakteristiky vztahu dvou proměnných jako analogii pro průměr nebo směrodatnou odchylku a popis chování jedné proměnné. Korelační analýza zkoumá vztahy kvantitativních a kategorických proměnných pomocí měr závislosti a grafů. Korelace udává míru stupně asociace dvou proměnných. Je zároveň velice důležité rozlišovat mezi korelací a kauzalitou, tedy přímým vztahem mezi příčinou a následkem. Podle Hendla (2012) totiž korelace sama o sobě ne-

znamená důkaz příčinného vztahu, tedy toho, že změny jedné proměnné skutečně působí na změny druhé proměnné.

Další využitou analytickou metodou byla regresní analýza. Podle Hindlse (2007) je regresní analýza soubor statistických metod, které se zabývají zkoumáním a hodnocením souvislostí mezi dvěma nebo více statistickými znaky. Cílem regresní analýzy je proniknutí do podstaty sledovaných jevů dané oblasti a přiblížení se k příčinným souvislostem, tj. souvislostem, kdy výskyt určitého jevu vyvolává existenci jevu dalšího. Regresní analýza zahrnuje několik variant, jako je lineární, vícenásobná lineární a nelineární. Nejběžnější modely jsou jednoduché lineární a vícenásobné lineární. Nelineární regresní analýza se běžně používá pro složitější soubory dat, ve kterých závislé a nezávislé proměnné vykazují nelineární vztah.

Pro všechny výpočty provedené v předkládaném výzkumu byl využit statistický software Statgraphics Centurion XVIII.

#### **Výsledky**

První výzkumná otázka byla zaměřena na zjištění, zda mají hráči dánské 3F Superligy na svém kontě v průměru vyšší počet vstřelených branek v porovnání s hráči české Fortuna:Ligy. Herní charakteristiky se ve fotbalovém prostředí obecně dělí na ofenzivní a defenzivní charakteristiky. K zjištění odpovědi na tuto výzkumnou otázku byly pro obě fotbalové ligy získány níže uvedené ofenzivní herní charakteristiky (Tabulka 1).

Dle vypočtených hodnot v Tabulce 1 je možné konstatovat, že většina součtů vybraných ofenzivních herních charakteristik dosahuje vyšších absolutních hodnot v případě české Fortuna:Ligy. Vyšší absolutní hodnoty jsou však také způsobeny vyšším počtem týmů a zároveň i hráčů v české Fortuna:Lize. Vyšších průměrných hodnot na druhé straně, ale dosahují ofenzivní herní charakteristiky v případě dánské 3F Superligy. Hráči v dánské 3F Superlize mají v průměru více celkových střel a také více střel na branku. Obdobnou situaci lze spatřit i v případě vstřelených gólů a asistencí. Modus u vybraných ofenzivních charakteristik je v tomto případě roven nule. To znamená, že jak v případě české Fortuna:Ligy tak i v případě dánské 3F Superligy nejčastěji hráči nedosahují za sezonu žádných zmíněných charakteristik.

Druhá výzkumná otázka byla zaměřena na zjištění, zda je česká Fortuna:Liga v porovnání s dánskou 3F Superligou více souborová. K zjištění odpovědi na tuto výzkumnou otázku byly pro obě fotbalové ligy získány níže uvedené defenzivní herní charakteristiky (Tabulka 2).



**Tabulka 1**

*Ofenzivní herní charakteristiky české Fortuna:Ligy a dánské 3F Superligy*

Herní charakteristika	Metrika	Fortuna:Liga	3F Superliga
<b>Góly</b>	Součet	3 201	3 150
	Průměr	1,61	1,82
	Modus	0	0
	Medián	1	1
	Var. rozpětí	29	29
<b>Asistence</b>	Součet	2 328	2 291
	Průměr	1,17	1,33
	Modus	0	0
	Medián	0	1
	Var. rozpětí	13	15
<b>Střely</b>	Součet	28 225	25 522
	Průměr	14,22	14,78
	Modus	0	0
	Medián	9	7
	Var. rozpětí	107	146
<b>Střely na branku</b>	Součet	10 635	9 898
	Průměr	5,36	5,73
	Modus	0	0
	Medián	3	2
	Var. rozpětí	50	63

**Tabulka 2**

*Defenzivní herní charakteristiky české Fortuna:Ligy a dánské 3F Superligy*

Herní charakteristika	Metrika	Fortuna:Liga	3F Superliga
<b>Souboje</b>	Absolutní	435 860	356 339
	Průměr	219,58	206,33
	Modus	0	0
	Medián	198	169
	Var. rozpětí	1 023	1 134
<b>Přestupky</b>	Absolutní	35 427	23 676
	Průměr	17,85	13,71
	Modus	0	0
	Medián	15	10
	Var. rozpětí	89	89
<b>Žluté karty</b>	Absolutní	4 838	4 062
	Průměr	2,44	2,35
	Modus	0	0
	Medián	2	2
	Var. rozpětí	14	14
<b>Červené karty</b>	Absolutní	213	171
	Průměr	0,11	0,10
	Modus	0	0
	Medián	0	0
	Var. rozpětí	2	2

Dle vypočtených hodnot v Tabulce 2 lze konstatovat, že hráči české Fortuna:Ligy dosáhli ve sledovaném období vyššího počtu soubojů, přestupků, žlutých i červených karet, jak v absolutní výši, tak také v průměrné výši. Vyšší absolutní hodnoty jsou, ale jako v případě první výzkumné otázky způsobeny vyšším počtem týmů a zároveň i hráčů v české Fortuna:Lize. V případě průměrných hodnot dosahuje česká Fortuna:Liga také většího průměrného množství podstoupených soubojů na zápas, většího průměrného množství přestupků a žlutých a červených karet.

Třetí výzkumná otázka se zaměřila na zjištění, ve které ze zkoumaných fotbalových soutěží působí větší počet zahraničních hráčů. K zjištění této výzkumné otázky byly pro obě fotbalové ligy od společnosti InStat získány údaje o počtu zahraničních hráčů, a to jak hráčů v poli, tak také brankářů.

Počet zahraničních hráčů v obou fotbalových ligách je zachycen na Obrázku 1. Z Obrázku 1 lze vyčíst, že větší náklonost k zahraničním hráčům, a to jak k hráčům v poli, tak také k brankářům je v dánské 3F Superlige. Na Obrázku 1 lze spatřit, že ve všech hodnocených sezonách je počet zahraničních hráčů v dánské 3F Superlige vyšší než v české Fortuna:Lize i přes to, že v dánské 3F Superlige je celkově méně týmů a zároveň tedy i menší počet hráčů.

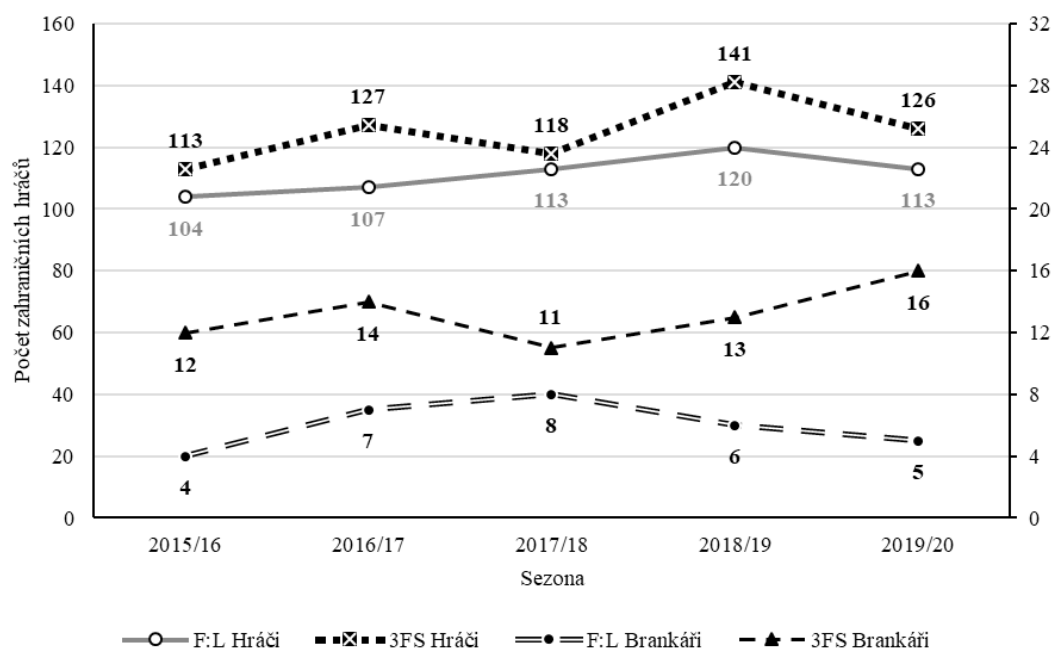
Čtvrtá výzkumná otázka se zaměřila na zjištění, ve které ze zkoumaných fotbalových soutěží je větší průměrná návštěvnost diváků. A zda závisí bodový zisk klubu na návštěvnosti. K zjištění této výzkumné otázky byly pro obě fotbalové ligy z databáze Bold (Bold, 2021) a Fortuna:Ligy (Fortuna:Liga, 2021) zjištěny údaje o průměrném počtu diváků v jednotlivých sezonách.

Mezi nejvíce navštěvované kluby v průběhu sledovaných sezon patřily v případě české Fortuna:Ligy v celkovém součtu domácích i venkovních zápasů AC Sparta Praha a SK Slavia Praha a v případě dánské 3F Superligy FC Copenhagen. V případě domácích zápasů byla v případě české Fortuna:Ligy zaznamenána největší návštěvnost s výjimkou sezony 2015/16 u klubu SK Slavia Praha. V případě venkovních zápasů byla zaznamenána největší návštěvnost u AC Sparta Praha. V případě domácích zápasů byla v případě dánské 3F Superligy zaznamenána největší návštěvnost u FC Copenhagen. V případě venkovních zápasů byla zaznamenána největší návštěvnost s výjimkou sezony 2015/16 u klubu Brøndby IF.

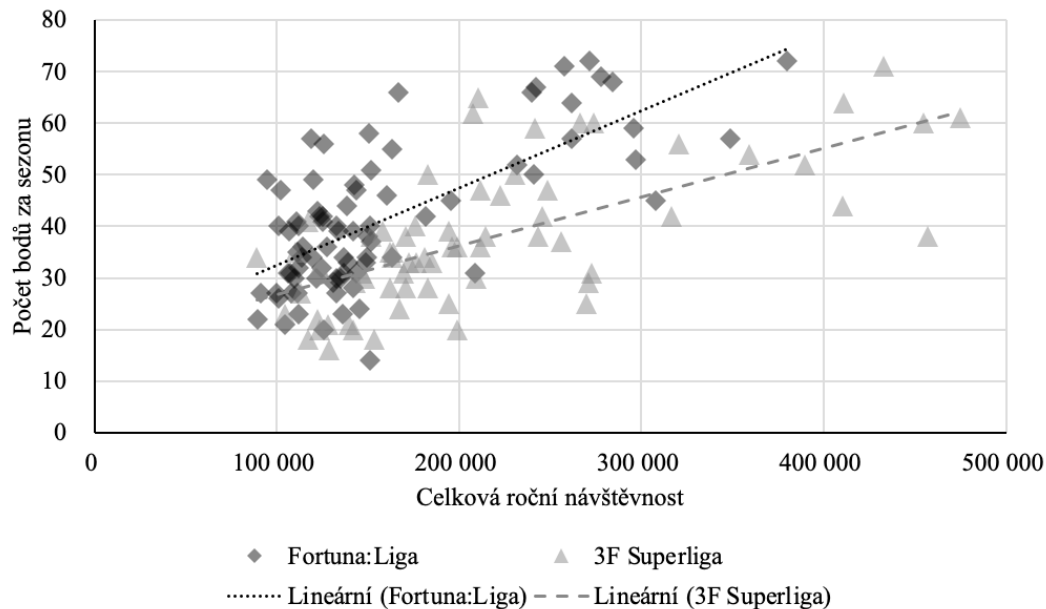
Mimo to byl zkoumán vztah mezi celkovou roční návštěvností a bodovým ziskem na konci sezony. Byla prokázána pozitivní středně silná závislost bodového zisku na celkovém počtu fanoušků na stadionu pro klub. Jak je znázorněno na Obrázku 2, klubům v české Fortuna:Lize s růstem celkové návštěvnosti rostl celkový bodový zisk na konci sezony rychleji než klubům v dánské 3F Superlige.

**Obrázek 1**

Počet zahraničních hráčů ve Fortuna:Lize a 3F Superlize v jednotlivých sezonách

**Obrázek 2**

Návštěvnost ve Fortuna:Lize a 3F Superlize ve vztahu k bodovému zisku klubu



Pátá výzkumná otázka zkoumala závislost tržní hodnoty hráčů v poli na jeho sportovní výkonnosti. Zároveň si výzkum kladl za cíl zjistit, zda roste tržní hodnota hráčů v poli dánské 3F Superligy s rostoucí sportovní výkonností rychleji v porovnání s hráči české Fortuna:Ligy. Ve fotbale jsou tržní hodnoty hráčů základním aktivem klubů, a to jak z hlediska herní, tak i z hlediska finanční výkonnosti. Tržní hodnota hráče odráží peněžní hodnotu, kterou zaměstnávající klub přisuzuje výkonu hráče. Tržní hodnoty fotbalových hráčů se proto obvykle obtížně hodnotí a jsou subjektivní. V rámci této výzkumné otázky byla sportovní výkonnost hráče v poli posuzovaná pomocí InStat indexu (InStat, 2020). Výzkum byl proveden souhrnně pro všechny fotbalové týmy z obou hodnocených soutěží za celé zkoumané období.

V předloženém výzkumu byla u hráčů v poli provedena korelační analýza, která potvrdila závislost mezi výkonností hráče v poli a jeho tržní hodnotou. Provedeným výzkumem byla zjištěna pozitivní středně silná korelace sportovní výkonnosti a tržní hodnoty hráčů. Hodnoty korelačních koeficientů jsou uvedeny v Tabulce 3.

Vyšší hodnota korelačního koeficientu byla zjištěna u hráčů v poli české Fortuna:Ligy nižší pak u hráčů v poli dánské 3F Superligy. Kromě korelačních koeficientů jsou v Tabulce 3 uvedeny také údaje o kvalitě vybraného modelu regresní analýzy a rovnice regresní přímky. U pozorovaných soutěží dosahoval nejvyšší hodnoty kvality regresního modelu model exponenciální. Z Tabulky 3 vyplývá, že jedna třetina dat přesně odpovídá zvolenému modelu. U hráčů české nejvyšší soutěže je tento poměr necelých 40 %. Regresní analýza pro hráče Fortuna:Ligy neodhalila statistickou významnost pro koeficient regresní přímky. Regresní analýza pro hráče 3F Superligy na druhé straně odhalila statistickou významnost pro koeficient regresní přímky.

Regresní analýza závislosti tržní hodnoty hráčů v poli na hodnotě InStat indexu všech hráčů v poli u obou analyzovaných soutěží za celé hodnocené období je znázorněna na Obrázku 3. Z Obrázku 3 je patrné, že regresní křivka hráčů české Fortuna:Ligy stoupá rychleji v porovnání s křivkou hráčů dánské 3F Superligy. Na základě toho lze konstatovat, že hráči v poli

české Fortuna:Ligy roste jeho tržní hodnota v závislosti na herních výkonech rychleji než hráči dánské 3F Superligy. Jedním z důvodů rychlejšího růstu tržní hodnoty hráčů v poli české Fortuna:Ligy oproti hráčům dánské 3F Superligy je skutečnost, že hráči dánské 3F Superligy mají v průměru vyšší tržní hodnotu než hráči české Fortuna:Ligy.

## Diskuze

V první části výzkumu byla provedena statistická charakteristika vybraných fotbalových soutěží, která byla zaměřena zejména na porovnání ofenzivních a defenzivních herních metrik. Z provedeného výzkumu vyplynulo, že v české Fortuna:Lize padlo ve sledovaném období i vzhledem k vyššímu počtu účastníků více branek. Na základě průměrných hodnot lze, ale na druhé straně konstatovat, že ve sledovaném období byla na vstřelené branky, střely na branku a střely bohatší dánská 3F Superliga.

Dánský fotbal je v České republice vnímán jako soubojový a fyzicky náročný. Větším průměrným množstvím podstoupených soubojů na zápas, ale disponuje na základě provedeného výzkumu česká Fortuna:Liga. Hráči české Fortuna:Ligy podstoupili více soubojů, častěji se provinili proti pravidlům a obdrželi větší množství žlutých a červených karet, jak v absolutních, tak také v průměrných hodnotách.

Lze tak konstatovat, že dánská 3F Superliga vychází z provedeného srovnání jako méně soubojová, kreativnější a pro diváka atraktivnější z pohledu většího množství vstřelených branek. Spolu s nižším množstvím účastníků dánské ligové soutěže oproti české Fortuna:Lize je na tamní kluby kladen větší tlak z pohledu konkurenceschopnosti.

Na základě provedené analýzy lze také konstatovat, že v dánské 3F Superlize se i přes nižší počet týmů vyskytovalo více zahraničních hráčů. Zastoupení cizinců mezi hráči v poli bylo okolo 36 % v dánské 3F Superlize a přibližně 28 % v české Fortuna:Lize. Mezi brankáři bylo v dánské 3F Superlize 47 % cizinců a v české Fortuna:Lize 16,67 % cizinců. Počet cizinců v dánské 3F Superlize odpovídá i hodnotám z provedeného výzkumu Maderera et al. (2014), který uváděl, že v prů-

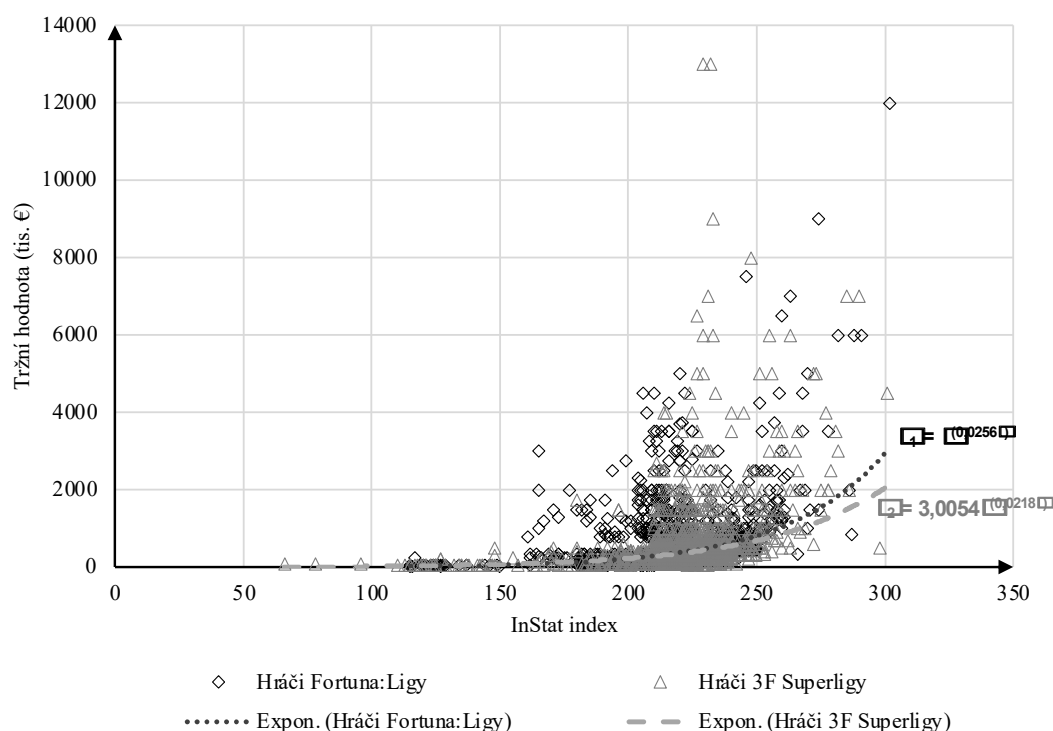
## Tabulka 3

Korelační koeficienty pro hráče v poli v celém hodnoceném období

Liga	Absolutní četnost	Pearsonův korelační koeficient	Model fit R-squared	Rovnice regresní přímky
F:L hráči v poli	1 985	0,63	39,34 %	$y = e^{(0,0256x)}$
3FS hráči v poli	1 727	0,52	26,82 %	$y = 3,0054e^{(0,0218x)}$
Všichni hráči v poli	3 712	0,58	33,65 %	$y = 1,9691e^{(0,0238x)}$

**Obrázek 3**

Regresní analýza závislosti tržní hodnoty na hodnotě InStat indexu všech hráčů v poli v obou analyzovaných soutěžích za celé hodnocené období



měru tvoří zahraniční hráči kolem 43 % všech hráčů. Po rozsudku Bosman v roce 1995 se kulturní rozmanitost profesionálních fotbalových týmů v Evropě značně zvýšila. Rozsudek belgického fotbalisty Jean-Mark Bosmana zásadně změnil přestupní pravidla ve fotbale. Tento rozsudek umožňuje fotbalistům v Evropské unii po vypršení smlouvy s fotbalovým klubem zdarma odejít do jiného fotbalového klubu. Rozsudek a jeho následky tak umožnily fotbalistům volnější pohyb mezi kluby v Evropě (Binder a Findlay, 2012). Nábor hráčů bez ohledu na jejich národnost umožňuje fotbalovým klubům využívat globální fond talentů a kombinovat specifické přednosti jednotlivců s různým kulturním zázemím. Lze konstatovat, že v současné době není trh s fotbalovými hráči regulován a je tak podpořena mezinárodní mobilita. Tento názor podporuje také Frick (2009).

Provedený výzkum rovněž prokázal pozitivní středně silnou závislost bodového zisku na celkovém počtu fanoušků na stadionu pro klub. K podobnému závěru také došel ve svém výzkumu Glass (2003), podle počet diváků koreluje s procentem vítězných zápasů domácích týmů.

Dále bylo ve výzkumu důležité zvážit, jak tržní hodnota hráče souvisí s výkonností tohoto hráče. U obou analyzovaných fotbalových soutěží byla prokázána středně silná korelace tržní hodnoty hráčů v poli a hodnoty InStat indexu. K podobnému závěru došli

i ve svém výzkumu He et al. (2015) a také Müller et al. (2017). Při aplikaci regresní analýzy nebyl vzhledem ke specifčnosti dat použit jednoduchý model lineární regrese. Vyšší kvality dosahoval dle hodnot koeficientu determinace exponenciální regresní model. Tento model odpovídá trendu navyšování tržních hodnot „rozdílových hráčů“, kteří svými sportovními výkony dokážou rozhodovat zápasy. U těchto hráčů je nárůst tržní hodnoty nelineární. Zvyšování tržní hodnoty nejlepších hráčů „táhne“ tržní hodnoty ostatních hráčů vzhůru. Ze souhrnné regresní analýzy lze říct, že tržní hodnota hráčů v poli české nejvyšší soutěže roste se zvyšováním hodnoty InStat indexu rychleji než u hráčů v poli dánské nejvyšší soutěže. Vedle sportovní složky mají na tržní hodnotu profesionálního fotbalového hráče vliv i další faktory, které se podílejí na tržní hodnotě hráčů (např. zkušenosti na evropských či světových šampionátech nebo jejich charakterové vlastnosti). K tomuto názoru se ve svém výzkumu přiklání také např. Majewski (2016).

**Závěry**

Cílem předkládaného výzkumu bylo porovnat dvě evropské fotbalové soutěže s vysokým potenciálem podobnosti pomocí statistické datové analýzy. Hodnocené fotbalové soutěže byly vybrány na základě podobných herních modelů, finančních rámců zúčastněných klubů



a také na základě jejich herních stylů. Ve výzkumu byly použity především metody popisné statistiky a také metody korelační a regresní analýzy. Každá z využitých metod umožnila analyzovat prostředí obou vybraných evropských fotbalových soutěží z jiného úhlu pohledu a zaměřovala se na specifické cíle.

Další možností budoucího výzkumu je obdobné srovnání nejvyšších fotbalových soutěží Evropy, nebo hlubší analýza jednotlivých aspektů ligových soutěží či jednotlivých fotbalových klubů a hráčů.

## Dedikace

Vznik příspěvku byl podpořen grantem Technické univerzity v Liberci č. 21456 s názvem „Faktory ovlivňující ekonomickou výkonnost evropských sportovních podniků“ (SGS-2022-1017).

## Reference

- 3F Superliga. (2021). *Superliga.dk: Performance Center*. <https://www.superliga.dk/performance-center-superliga-2020-2021>
- Binder, J. J., & Findlay, M. (2012). The effects of the Bosman ruling on national and club teams in Europe. *Journal of Sports Economics*, 13(2), 107–129. <https://doi.org/10.1177/1527002511400278>
- Bold. (2021). *bold.dk: 3F Superliga – Tilskuere*. <https://www.bold.dk/fodbold/danmark/3f-superliga/2015-2016/>
- Broich, H., Mester, J., Seifriz, F., & Yue, Z. (2014). Statistical analysis for the first Bundesliga in the current soccer season. *Progress in Applied Mathematics*, 7(2), 1–8. <https://doi.org/10.3968/4886>
- 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>
- Delgado-Bordonau, J. L., Doménech-Monforte, C., & Guzmán, J. (2013). Offensive and defensive team performance: Relation to successful and unsuccessful participation in the 2010 Soccer World Cup. *Journal of Human Sport and Exercise*, 8(4), 894–904. <https://doi.org/10.4100/jhse.2013.84.02>
- dos Santos, C. A., Dani, A. C., & Hein, N. (2016). Relationship between the Brazilian soccer confederation rankings and the economical-financial indicators of soccer teams. *Podium-sport Leisure and Tourism Review*, 5(3), 41–59. <http://dx.doi.org/10.5585/podium.v5i3.161>
- Fortuna:Liga. (2021). *FortunaLiga.cz: Statistika*. <https://www.fortunaliga.cz/statistika?unit=1&parameter=1>
- Frick, B. (2009). Globalization and factor mobility: The impact of the „Bosman-ruling“ on player migration in professional soccer. *Journal of Sports Economics*, 10(1), 88–106. <https://doi.org/10.1177/1527002508327399>
- 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>
- Glass, D. B. (2003). Fair-weather fans: The correlation between attendance and winning percentage. *The Baseball Research Journal*, 21, 81–84.
- Harper, J. (2021, March 5). *Data experts are becoming football's best signings*. BBC News. <https://www.bbc.com/news/business-56164159>
- He, M., Cachucho, R., & Knobbe, A., & He, M. (2015). Football player's performance and market value. In J. Davis, J. Van Haaren, & A. Zimmermann (Eds.), *Proceedings of the 2nd workshop of sports analytics, European conference on machine learning and principles and practice of knowledge discovery in databases (ECML, PKDD)*. (pp. 87–95). [https://www.researchgate.net/publication/321623604\\_Football\\_player%27s\\_performance\\_and\\_market\\_value](https://www.researchgate.net/publication/321623604_Football_player%27s_performance_and_market_value)
- Hendl, J. (2012). *Přehled statistických metod: Analýza a meta-analýza dat* (4th ed.). Portál.
- Hindls, R. (2007). *Statistika pro ekonomy*. (8th ed.). Professional Publishing.
- InStat. (2020, September). *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 football transfer market. *Concurrency and Computation Practice and Experience*, 33(3), e5353. <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>
- Maderer, D., Holtbrügge, D., & Schuster, T. (2014). Professional football squads as multicultural teams: Cultural diversity, intercultural experience, and team performance. *International Journal of Cross Cultural Management*, 14(2), 215–238. <https://doi.org/10.1177/1470595813510710>
- Majewski, S. (2016). Identification of factors determining market value of the most valuable football players. *Journal of Management and Business Administration*, 24(3), 91–104. <https://doi.org/10.7206/jmba.ce.2450-7814.177>
- 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>
- Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, 263(2), 611–624. <https://doi.org/10.1016/j.ejor.2017.05.005>
- 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), Article 10. <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.
- Peters, D. (2011). *Winning percentage and attendance in the NHL*. St. John Fisher College.

- 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 soccer game 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.
- Transfermarkt. (2021). *Transfermarkt.com: Intern*. <https://www.transfermarkt.com/intern/stellenangebote>
- 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>

## Sports data analysis: Comparison of the Czech and Danish top football competitions

**Background:** Football is one of the most popular sports in the world and the use of statistical models to analyse data is of great interest to coaches, sports managers, fans, journalists and, of course, the athletes themselves. In recent years, football clubs have very often used the services of companies that analyse sports data. With the help of data analysis, it is possible to find out information that cannot be ascertained simply by observing the game. It is no wonder that there is currently a great deal of interest in sports data analysis, especially among football clubs, which make important strategic decisions based on them. **Objective:** Use statistical analysis to compare two professional European football competitions and identify their main similarities and differences. **Methods:** Methods of statistical data analysis were used to verify the research questions formulated in this research. The database was created based on variables obtained from InStat, which analyses the performance of athletes and sports teams. Statgraphics Centurion XVIII statistical software was used for the calculations. **Results:** Danish football is based on a comparison with Czech football as less competitive, more creative and more attractive for the spectator in terms of a larger number of goals scored. Danish football has a higher proportion of foreigners among the players. In both analysed competitions, a positive moderate dependence of point gain on the total number of fans at the stadium for the club was demonstrated. With the growth of total attendance, the Czech clubs grew their overall points at the end of the season faster than the Danish clubs. In both analysed competitions, a medium-strong correlation between the market value of players in the field and their performance was demonstrated. For a Czech player in the field, his market value grows faster depending on the game performance than for a player in a Danish competition. **Conclusions:** The research compared two European football competitions with a high potential for similarity using statistical data analysis. Data analysis in sports can be recommended in decision-making processes, mainly due to its economic and sports benefits.

**Keywords:** football; correlation analysis; regression analysis; market value of football players; sports analysis