Premier liga

Filip Grčar

17. februar 2018

1 Uvod

Tako kot mnogo drugih športnih panog tudi pri nogometu obstaja precej nenapisanih pravil, ki se pogosto v praksi izkažejo za resnične, ali pa pravil, ki jih vsi poznajo, a le redko kdo pozna njihovo statistično ozadje. Če vsaj malo spremljaš nogomet, zagotovo poznaš pravilo gola doseženega na gostovanju, ki pravi, da v primeru, da sta ekipi po dveh medsebojnih tekmah izenačeni, napreduje tisto moštvo, ki je doseglo več golov v gosteh. Toda, ali drži, da domače moštvo lažje zmaga? Pravi naslov za ugotavljanje tega in podobnih pravil so statistični podatki, iz katerih, z nekaj znanja, pridemo do marsikaterih ugotovitev. Na omenjeno vprašanje v nadaljevanju sledi odgovor, ki temelji na statističnih podatkih angleške Premier lige, pogledali pa si bomo še nekatere značilnosti angleškega nogometa in na podlagi statističnih podatkov kakšno stvar tudi skušali predvideti:

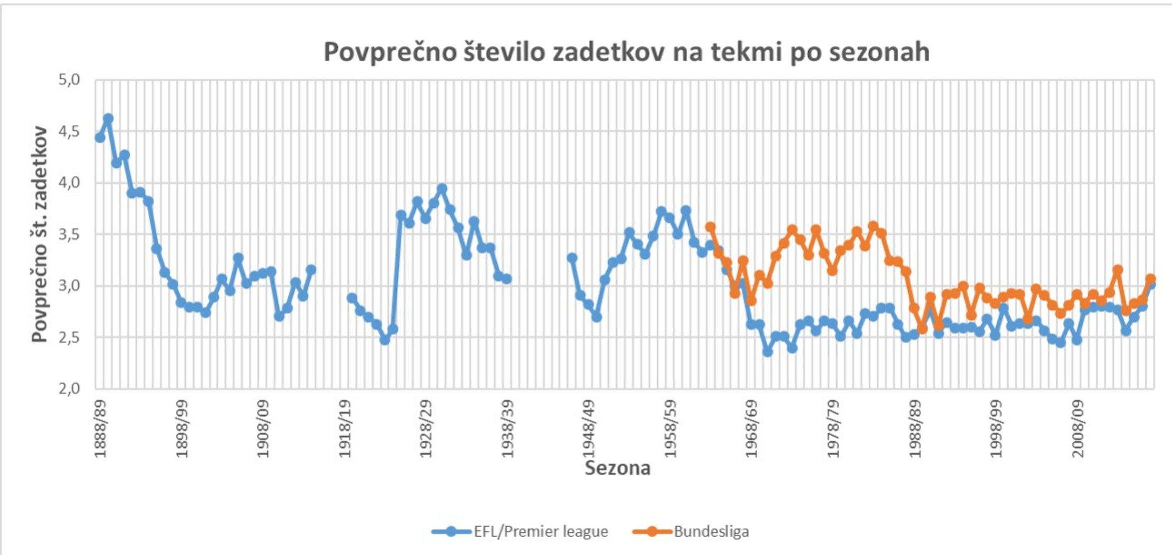
- povprečje zadetkov na tekmo v Premier ligi skozi zgodovino ter primerjava z nemško Bundesligo (2. razdelek),
- zahtevnost osvajanja Premier lige, kjer je prikazan povprečen izgled lestvice ob koncu sezone, osredotočili pa se bomo predvsem na prvo mesto (3. razdelek),
- preverjanje naslednjih domnev (4. razdelek):
    1. ekipa s 3 zadetki zanesljivo zmaga,
    2. domača ekipa večinoma zmaga,
    3. ekipa s prednostjo v posesti žoge večinoma zmaga,
    4. boljše ekipe imajo večjo podporo s tribun.
- vpliv doseženih zadetkov na končno število točk (5. razdelek),
- vpliv strelov na gol na število doseženih zadetkov (6. razdelek),
- napoved rezultata tekme večnih rivalov Manchester United-a in Liverpool-a (7. razdelek).

Cilj seminarske naloge je torej analiza nekaterih podatkov Premier lige in ugotavljanje njihove povezanosti ter možna uporaba ugotovitev za napovedovanje rezultatov v prihodnosti.

Ker je Premier liga najbolj priljubljena liga na svetu, obstaja že precej analiz le-te. Večina se ukvarja z napovedovanjem izidov na podlagi različnih koeficientov obrambe, napada oz. verjetnosti zmage posameznih moštev (glej [5]) ali pa z razvojem posameznih igralnih položajev (glej [6]). Sam sem se lotil analize različnih podatkov, ki sem jih zbral na več načinov. Večino podatkov sem pridobil z uradne spletne strani Premier lige (glej [3]). Nekaj podatkov je bilo takih, da so že povzemali povprečne vrednosti za daljše obdobje, nekaj pa sem jih moral pridobiti sam tako, da sem se osredotočil na lansko sezono (2016/17) in se sprehodil preko vseh tekem in z vsake zbral podatke, ki sem jih v nadaljevanju uporabil za dokazovanje določenih domnev. Nekatere podatke sem pridobil s spletne strani Wordfootball (glej [4]), ki vsebuje podatke različnih lig, kjer sem pridobil tudi podatke za primerjavo Premier lige z nemško Bundesligo. Ker je bilo podatkov precej, sem za obdelavo uporabil Microsoft Excel.

2 Povprečje zadetkov na tekmo

Za začetek si poglejmo, kako se je gibalo povprečje zadetkov na posamezni tekmi skozi zgodovino, saj je to ena izmed najpomembnejših nogometnih prvin. Brez zadetkov ekipa pač ne more zmagati. Kot vidimo na grafu 1, je bilo na začetku elitne angleške nogometne lige (takrat še EFL oz. England Football League, Premier league od 1992 naprej) v povprečju doseženih več kot 4 zadetke na tekmo, potem pa se je v naslednjih letih zmanjšalo. Večji skok opazimo le v času med vojnama, od leta 1980 naprej pa se povprečje giblje okrog 2.5 do 3 zadetke na tekmo. Za primerjavo sem dodal nemško Bundesligo in kot lahko opazimo, je tam doseženih nekoliko več zadetkov. V nekaj letih pred 1990 opazimo padec povprečja zadetkov v nemški ligi, kar je obdobje pred padcem železne zavese in morda so se takrat Nemci začeli bolj zgledovati po angleškem nogometu. Opazimo tudi trend, da imata obe ligi v začetku višje povprečje, nato se zmanjša, kar pomeni, da je v igri vedno bolj pomemben postal tudi element obrambe.



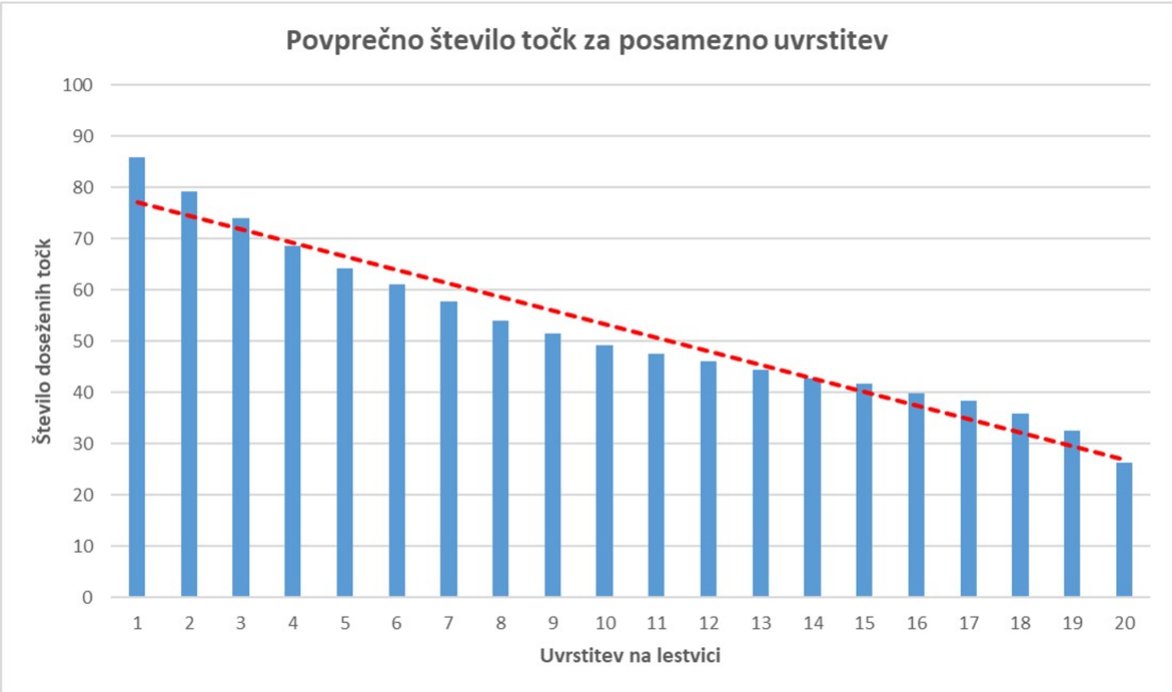
Slika 1: Graf povprečja zadetkov na posamezni tekmi po sezonah za angleško in nemško nogometno ligo

Povprečje ($\bar{x}$) in popravljen standardni odklon (s) izračunamo s spodnjima enačbama, glej [1, str. 28, 33]:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s = \sqrt{\frac{\sum_{i=1}^{n}(\bar{x}-x_i)^2}{n-1}}$$

in dobimo skupno povprečje vseh sezon angleške prve lige, ki znaša malenkost več kot 3 zadetke (3.01) s popravljenim standardnim odklonom 0.49 zadetka. Za nemško Bundesligo je povprečje malenkost višje, to je 3.07 zadetka na tekmo s popravljenim standardnim odklonom 0.28 zadetka. Iz izračunanih odklonov in grafa 1 vidimo, da je število zadetkov na tekmo v nemški ligi bolj konstantno, v zadnjih letih pa med ligama ni bistvene razlike. Razlog za to lahko najdemo v sodobnem nogometu. Stil igre na najvišjem nivoju ni več tako različen, a vseeno obstajajo določene značilnosti. Angleški nogomet je bolj direkten, kar pomeni dolge podaje čez sredino igrišča, veliko zračnih dvobojev, torej bolj fizičen, grob nogomet. Za nemški nogomet pa je značilna dobra organiziranost ekipe, agresivnost in morda je posledica tega večje število zadetkov.

3 Zahtevnost osvajanja Premier lige



Slika 2: Povprečno število točk za posamezno uvrstitev na lestvici

Na grafu 2 vidimo povprečje osvojenih točk za posamezno uvrstitev na lestvici Premier lige za zadnjih 22 (n) sezon, prej je bilo v ligi namreč več ekip in zato prejšnjih sezon ne moremo enakovredno primerjati. Za osvojitev prvega mesta so zmagovalne ekipe v povprečju dosegle skoraj 86 točk, kar pomeni, da so v povprečju na 38 tekmah dosegle 28 zmag in 2 remija oz. kakšno zmago manj in za vsako manj po 3 remije več. Minimalno število točk za prvo mesto je v sezoni 1996/1997 dosegla ekipa Manchester United-a, to je 75 točk, maksimalno pa v sezoni 2004/2005 z le enim samim porazom ekipa Chelsea-ja, to je 95 točk. Poglejmo si še interval zaupanja za točke ekipe, ki osvoji prvo mesto lestvice, s stopnjo tveganja 10% (α). Vzorčno povprečje ($\bar{x}$) je 85.8, standardni odklon s 5.3, t predstavlja Studentovo porazdelitev in μ pričakovano vrednost oz. $E(X)$. Tako interval zaupanja izračunamo po spodnjem postopku, glej [1, str. 282]:

$$\Delta = \frac{s \cdot t_{1-\alpha/2}(n-1)}{\sqrt{n}}, \qquad \Delta = \frac{5.3 \cdot 1.72}{\sqrt{22}} = 1.94,$$

$$\bar{x} - \Delta \leq \mu \leq \bar{x} + \Delta, \qquad \mu \in [83.9, 87.7].$$

S stopnjo zaupanja 90% lahko torej trdimo, da je povprečje točk za prvo mesto na intervalu $[83.9, 87.7]$. To pomeni, da je $P(\bar{x} - \Delta \leq \mu \leq \bar{x} + \Delta) = 1 - \alpha$. Če za stopnjo tveganja privzamemo 5%, se interval zaupanja razširi na $[83.5, 88.2]$, saj za širši interval lahko z večjo verjetnostjo (95%) trdimo, da je povprečje na tem intervalu.
Izračunajmo še korelacijo med osvojenim mestom in povprečnim številom doseženih točk v sezoni. Korelacijski koeficient izračunamo po spodnji enačbi, glej [1, str. 465]:

$$r_{X,Y} = \frac{k(X,Y)}{s_X s_Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} = -0.974.$$

Kot vidimo je absolutna vrednost korelacijskega koeficienta blizu 1, kar pomeni močno linearno odvisnost med spremenljivkama. Vrednost je pričakovana, saj na grafu 2 vidimo relativno majhna odstopanja vrednosti od regresijske premice. Določimo še enačbo le-te, ki jo dobimo po spodnji enačb, glej [1, str. 445]:

$$y = \bar{y} + \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}(x - \bar{x}),$$

kar nam da $\mathbf{y = -2.63 \cdot x + 79.70.}$

4 Domneve

V tem poglavju si bomo pogledali nekaj domnev s področja nogometa in jih poskušali dokazati oz. ovreči na podatkih Premier lige.

1. domneva: ekipa, ki na tekmi doseže vsaj 3 zadetke, zanesljivo zmaga

Vsak navijač trepeta za zmago, če njegovo moštvo vodi z le 1:0. Tudi 2:0 še ni dovolj zanesljiv rezultat, da si navijači lahko oddahnejo. So torej trije goli dovolj? V lanski sezoni (2016/17) je bilo 135 tekem, na katerih je vsaj ena ekipa dosegla 3 ali več zadetkov in 127 tekem se je končalo z zmago ekipe s 3 ali več zadetki, na 8 tekmah pa kljub 3 zadetkom ene ekipe to ni bilo dovolj za zmago, saj jih je nasprotna ekipa dosegla enako ali celo več. Za stopnjo tveganja (α) privzamemo 10%, za mejo, kjer še trdimo, da 3 zadetki pomenijo zanesljivo zmago (p_0)

pa vzamemo 90%. Kvantil z_α v enačbi pomeni, da gre za normalno porazdelitev. Opravimo torej test, glej [1, str. 369]:

$$n = 135, \qquad k = 127, \qquad H_0 : p = p_0, \qquad H_\alpha : p > p_0,$$
$$\mathrm{TS} = \sqrt{\frac{n}{p_0(1-p_0)}}\left(\frac{k}{n} - p_0\right) = 1.576, \qquad K_\alpha = (z_\alpha, \infty) = (1.28, \infty),$$
$$P\text{-vrednost} = 0.0575.$$

Kot vidimo pri zgornjem postopku se testna statistika nahaja v kritičnem območju K_α, kar pomeni, da ovržemo ničelno domnevo H_0 proti alternativni domnevi H_α in torej lahko z 10% stopnjo tveganja trdimo, da ekipa, ki doseže vsaj 3 zadetke na tekmi zmaga v več kot 90% primerih. To potrjuje tudi P-vrednost, ki je manjša od α. P-vrednost je namreč najmanjša stopnja značilnosti, pri kateri še zavrnemo ničelno domnevo pri danih podatkih. Če zmanjšamo stopnjo tveganja na 5%, je kritično območje $K_\alpha = (1.645, \infty)$, in testna statistika se ne nahaja v tem območju. Ničelne domneve v tem primeru ne moremo zavrniti. Če za zanesljivost zmage (p_0) vzamemo 95% ali več, je testna statistika negativna in se tako zagotovo ne nahaja v kritičnem območju, kar pomeni, da zagotovo ne moremo trditi, da 3 zadetki prinašajo zmago v 95% ali več primerih.

2. domneva: domača ekipa večinoma zmaga (ne izgubi)

Slika 3: Kolač deleža zmag glede na teren

V nogometu je pravilo, da zadetek, dosežen v gosteh, velja več, kot zadetek dosežen doma. Pričakovano imajo domači igralci večjo podporo s tribun in so zato tudi bolj motivirani ter posledično lažje iztržijo dober rezultat, a kaj pravi statistika? Je torej res, da domača ekipa ponavadi zmaga? Na podlagi podatkov lanske sezone si poglejmo ali to res drži. V sezoni 2016/17 je bilo odigranih 380 tekem, na keterih je bilo 84 remijev, 109-krat so zmagali gostje in kar 187-krat domači. Glede na podatke vidimo, da je domača ekipa zmagala na 49.2% tekmah oz. na 71.3% tekmah ni izgubila. To vidimo tudi na grafu 3. Če delimo število domačih zmag s številom zmag gostov, dobimo kvocient 1.7. Torej je domača zmaga skoraj dvakrat verjetnejša od zmage gostov. Poskušajmo torej dokazati, da domače moštvo zanesljivo ne izgubi, za izračun pa uporabimo enaki predpostavki kot pri prejšnji domnevi, torej $p_0 = 90\%$ in $\alpha = 10\%$, glej [1, str. 369]:

$$n = 380, \qquad k = 271, \qquad H_0 : p = p_0, \qquad H_\alpha : p > p_0,$$
$$\mathrm{TS} = -12.14, \qquad K_\alpha = (1.28, \infty), \qquad P\text{-vrednost} \approx 1$$

Kot vidimo je testna statistika daleč od kritičnega območja in zato nikakor ne moremo trditi, da domače moštvo zanesljivo ne izgubi, kar pa je pričakovano, saj je delež vendarle prenizek (71%). Domneva, da domača ekipa zanesljivo ne izgubi je torej napačna, vendar domnevo, da domače moštvo večinoma ne izgubi, lahko potrdimo, saj 71% verjetnost vseeno potrdi dejstvo, da je domača ekipa nekoliko v prednosti. Za razlog, da dani odstotek ni še višji, lahko vzamemo dejstvo, da kvalitetnejše ekipe vseeno znajo unovčiti svojo premoč, ne glede na teren, če pa je tekma bolj izenačena, pa teren lahko nagne rezultat v korist domačega moštva.

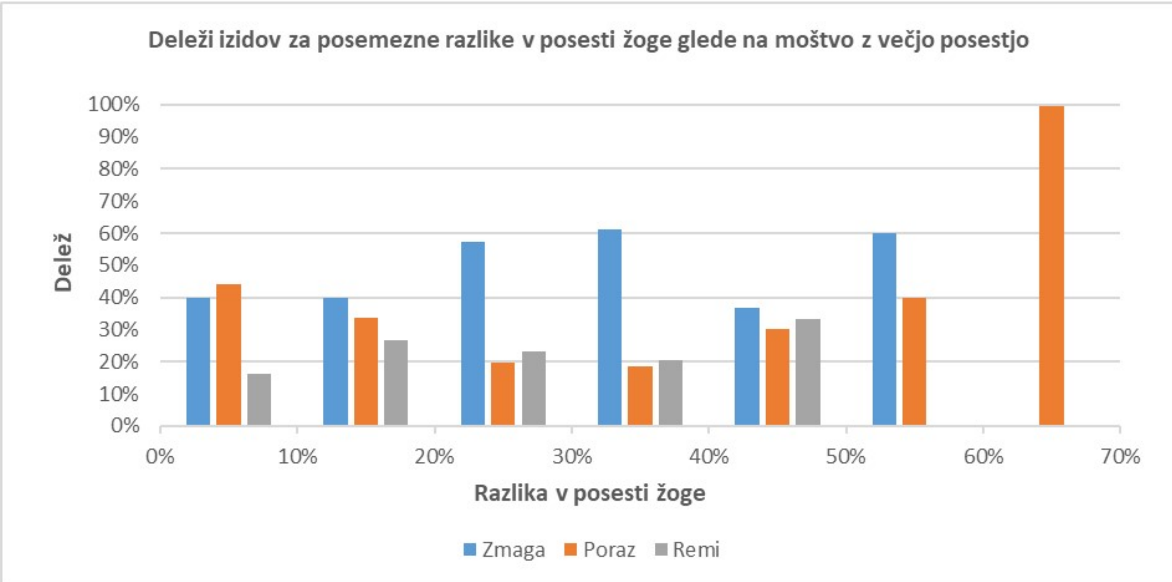
3. domneva: ekipa z večjo posestjo žoge večinoma zmaga (ne izgubi)

Posest žoge pri nogometu pomeni razmerje časa med ekipama, ko je imela posamezna ekipa žogo v svoji posesti. Pričakovano ima dominantnejša ekipa več časa žogo v svoji posesti in zato bi pričakovali, da le ta tudi zmaga, a v nogometu ni vse tako preprosto. Poglejmo si podatke iz lanske sezone (tabela 1).

izid\razlika	0 - 9.9%	10 - 19.9%	20 - 29.9%	30 - 39.9%	40 - 49.9%	50 - 59.9%	60 - 69.9%	$\sum$
zmaga	40	39	50	36	11	3	0	179
poraz	44	33	17	11	9	2	1	117
remi	16	26	20	12	10	0	0	84

Tabela 1: Predstavljene so vse tekme lanske sezone (2016/2017) in sicer glede na razliko v posesti med moštvoma in kakšen je bil rezulatat za moštvo z vodstvom v posesti (npr. razmerje posesti žoge med moštvom 1 in moštvom 2 je 54% : 46%, moštvo 1 pa je izgubilo, se dana tekma šteje v razdelek poraz, 0-9.9%). Iz vrstičnih vsot vidimo, da je moštvo z večjo posestjo zmagalo 179-krat (47.1%), 117-krat (30.8%) je prišlo do remija, 84-krat (22.1%) pa je ekipa z večjo posestjo izgubila.

Poglejmo še deleže izidov za posamezne razlike v posesti, graf 4.



Slika 4: Graf deležev izidov za posemezne razlike v posesti žoge glede na moštvo z večjo posestjo

Korelacijski koeficienti razlike v posesti glede na deleže zmag, porazov in remijev so:

$$\text{zmage: } r_{X,Y} = -0.363, \quad \text{porazi: } r_{X,Y} = 0.530, \quad \text{remiji: } r_{X,Y} = 0.592.$$

Ker absolutne vrednosti koeficientov niso blizu 1, pomeni, da med razliko v posesti in deležem izidov ni močne linearne zveze, to se vidi tudi na grafu 4, kjer težko narišemo regresijske premice, saj bi bila odstopanja vrednosti od premic prevelika. Pričakovali bi, da delež zmag naraste, remijev in porazov pa pade, a ni tako. Delež zmag nima izrazite rasti in se večinoma giblje med 40 in 60%. Delež remijev prav tako niha, večinoma med 10 in 30%. Delež porazov najprej z naraščanjem razlike posesti res upada, a nato začne naraščati. V praksi to pomeni, da moštvo, ki je v izrazito podrejenem položaju oz. igra "bunker", pogosto zmaga preko protinapadov. Poglejmo nekaj takih primerov npr. Burnley : Liverpool, ko je bilo razmerje v posesti žoge 19.6 : 80.4%, rezultat pa 2:0 (odigrano 20. 8. 2016), ali Everton : Manchester City, ko je bilo razmerje v posesti žoge 29.2 : 70.8%, rezulatat pa 4:0 (odigrano 14. 1. 2017), ali Crystal Palace : Hull (29.7 : 70.3%, rezultat pa 4:0, odigrano 14. 5. 2017). V danih primerih se izkaže za pravilno nepisano nogometno pravilo "kdor ne da, dobi".

Kot kaže večinoma ekipa z večjo posestjo res v prednosti, saj v 69% ne izgubi, a kot vidimo, so tudi ekipe, ki so imele mnogo večjo posest (npr. več kot 40% več), vseeno izgubile. Domnevo torej lahko delno potrdimo, a ne v celoti, ravno zaradi tistih tekem z veliko prednostjo poražene ekipe.

4. domneva: bolje uvrščene ekipe imajo večjo podporo s tribun

Chelsea	Tottenham	Man. City	Liverpool	Arsenal	Man. United	Everton
99.2%	98.8%	98.0%	98.0%	99.3%	99.4%	97.8%

Southampton	Bournemouth	West Bromwich	WHU	Leicester	Stoke
94.6%	97.5%	90.3%	95.0%	97.4%	91.2%

Crystal Palace	Swansea	Burnley	Watford	Hull	Middlesbrough	Sunderland
95.8%	99.4%	91.2%	93.1%	81.1%	86.8%	84.3%

Tabela 2: Povprečna obiskanost tekem, pri čemer so ekipe zapisane v vrstnem redu na lestvici

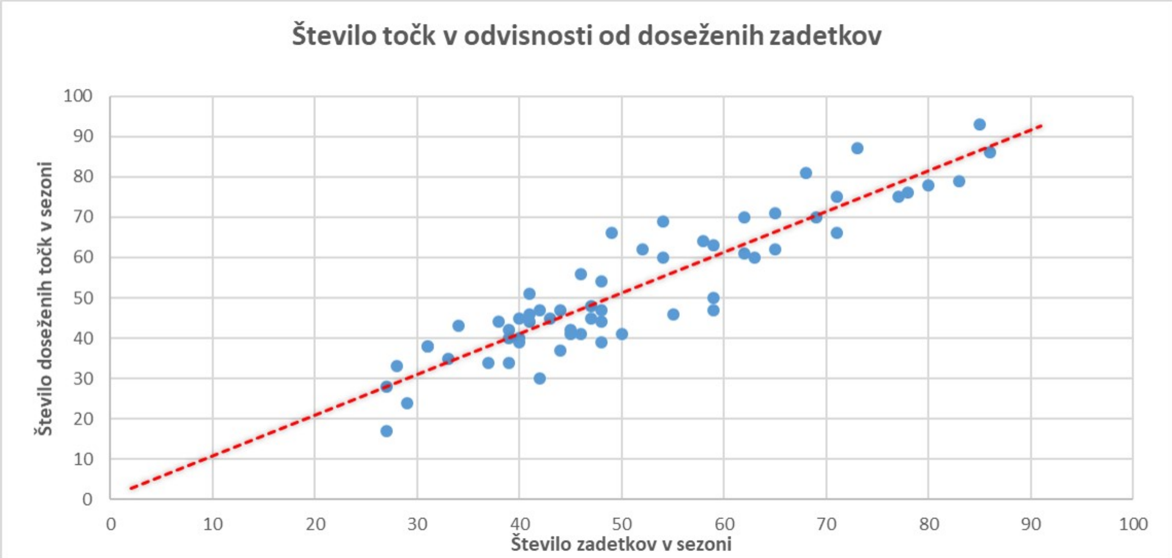
V tabeli 2 so predstavljene povprečne obiskanosti tekem za posamezno moštvo. Moštva pa so urejena po končni razvrstitvi v sezoni 2016/2017. Kot vidimo ni bistvene razlike v obiskanosti tekem med boljše in slabše uvrščenimi ekipami, zato domnevo zavrnemo. Kot kaže, so pravi nogometni navijači prava klapa - ljubezen do kluba je močnejša od slabih rezultatov.

5 Vpliv doseženih zadetkov na končno število točk

Kot pri drugih športih, tudi pri nogometu velja, da, kdor doseže več zadetkov na tekmi, zmaga. Pa v povprečju več doseženih zadetkov (X) prinese tudi več točk (Y)? To si bomo pogledali v nadaljevanju. Predvidevam, da bi marsikdo rekel: "Kje so pa tu prejeti zadetki?", kajti ekipe, ki prejmejo manj zadetkov, jih za zmago potrebujejo tudi manj zadeti. Vendar kot bomo videli v nadaljevanju, je tudi med doseženimi zadetki (in ne gol razliko) in končnim številom točk precej močna korelacija. Pa izračunajmo korelacijski koeficient, da se prepričamo, če to res drži:

$$r_{X,Y} = 0.917.$$

Kot vidimo je korelacijski koeficient zelo blizu 1, kar pomeni, da sta število doseženih zadetkov in končno število točk linearno povezani. Odvisnost je lepo razvidna tudi na sliki 5, ki prikazuje odvisnost končnega števila točk od doseženih zadetkov za zadnje tri sezone Premier lige (2014/15, 2015/16, 2016/17).



Slika 5: Graf končnega števila točk v sezoni v odvisnosti od doseženih zadetkov v sezoni

Izračunajmo še enačbo regresijske premice, po kateri lahko nato iz doseženih zadetkov predvidimo okvirno število točk za neko ekipo, enačba regresijske premice: **št. točk = 1.0082 · št. doseženih zadetkov + 0.7646**.

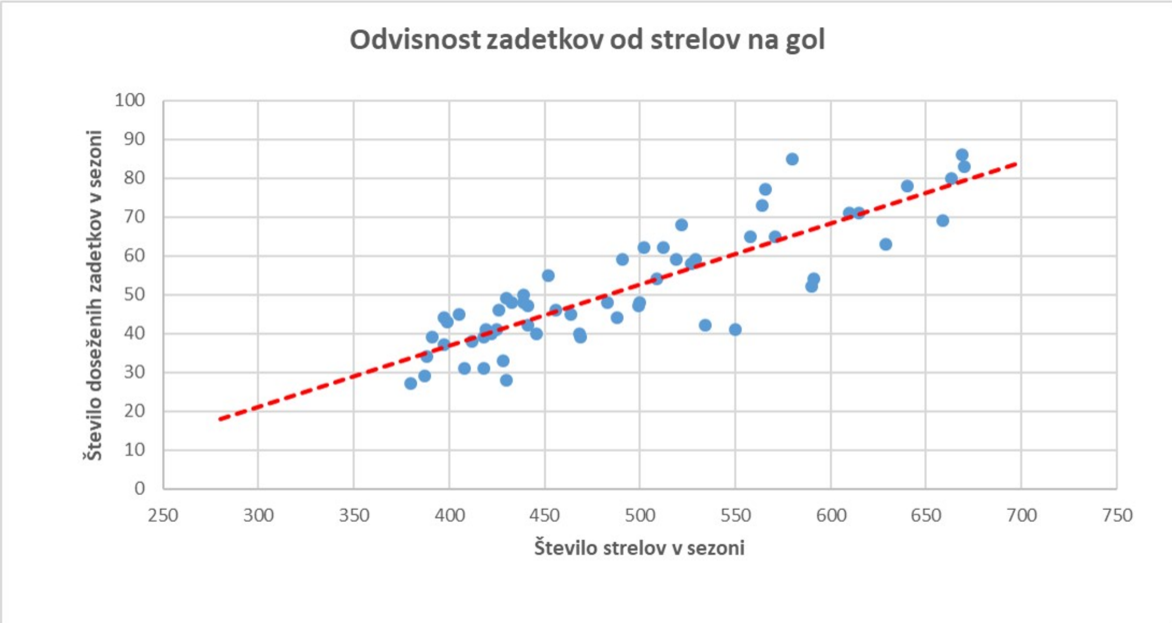
Npr. če je ekipa v sezoni dosegla v povprečju gol na tekmo (38 golov v sezoni), ji to po naši oceni prinaša 39 točk. Glede na povprečje točk za posemezno mesto na lestvici (graf 2), to prinaša skromno 17. mesto na lestvici, ki pa vseeno pomeni obstanek v ligi.

6 Vpliv strelov na gol na število doseženih zadetkov

Kot smo ugotovili, več zadetkov prinese več točk, za zadetke pa so potrebni streli na gol, saj, če ne poiskusiš, ne moreš zadeti. Pričakujemo torej, da kdor večkrat strelja, tudi doseže več zadetkov. Za začetek si spet poglejmo korelacijski koeficient med streli na gol (X) in številom doseženih zadetkov (Y):

$$r_{X,Y} = 0.861.$$

Korelacijski koeficient je blizu 1, torej gre za močno linearno odvisnost, sicer nekoliko manj kot zadetki in točke, a vseeno kar močno. Za lažjo predstavo je odvisnost predstavljena na sliki 6, ki prikazuje odvisnost števila zadetkov od strelov na gol za zadnje tri sezone Premier lige (2014/15, 2015/16, 2016/17).



Slika 6: Graf števila zadetkov v sezoni v odvisnosti od strelov na gol

Enačba regresijske premice: **št. zadetkov = 0.1578 · št. strelov na gol − 26,2319**.

Ker smo našli povezavo med številom strelov na gol in zadetki ter zadetki in točkami, lahko zaključimo, da obstaja tudi odvisnost med streli na gol in točkami. Iz tega lahko sklepamo, da boljše ekipe zaradi svoje kvalitete večkrat pridejo do strela oz. prave priložnosti, posledično dosežejo več zadetkov in osvojijo več točk.

7 Napoved rezultata tekme

Pogledali si bomo, kako verjeten je nek rezultat na tekmi med velikima rivaloma Manchester United-om in Liverpool-om. Ker smo prej ugotovili, da je domače moštvo v prednosti, tukaj predpostavimo, da se tekma igra na nevtralnem terenu z enakim številom navijačev obeh moštev. Na 50 tekmah med Manchester United-om in Liverpool-om je United zmagal 27-krat, Liverpool 13-krat, bilo pa je še 10 remijev. Manchester United je skupno dosegel 73 zadetkov, Livepool pa 56. Za lažjo napoved predpostavimo, da je verjetnost zadetka na tekmi porazdeljena po Poissonu, kar pomeni, da pričakovano število zadetkov na tekmi za United 1.46 (λ_1), za Liverpool pa 1.12 (λ_2). Da ne bomo le prepostavljali, si poglejmo, ali izbrana porazdelitev ustreza na primeru Manchester United-a. Za dokaz uporabimo test hi-kvadrat. Za začetek določimo ničelno in alternativno domnevo, ki sta

H_0 = število zadetkov je porazdeljeno po Poissonu,
H_1 = število zadetkov ni porazdeljeno po Poissonu ($H_1 = \overline{H_0}$).

Sledi primerjava dejanskih frekvenc za posamezno število zadetkov (f_d) in teoretičnih frekvenc (f_t), ki prikazujejo vrednosti, če bi bili zadetki porazdeljeni po Poissonu, podatki so predstavljeni za vzorec 50 tekem (glej tabelo 3).

št. zadetkov\frekvence	f_d	f_t
0	11	11.61
1	15	16.95
2	15	12.38
3	8	6.02
4	1	2.20

Tabela 3: Dejanske in teoretične frekvence zadetkov Manchester United-a na 50 tekmah

Izračunajmo testno statistiko, ki meri prilagojenost dejanskih frekvenc teoretičnim. Testna statistika se porazdeljuje po χ^2-porazdelitvi z $k-p-1$ prostostnimi stopnjami, kjer je k število razredov, v katere so porazdeljeni podatki, p pa število podatkov pridobljenih iz podatkov, katerih frekvence računamo. V našem primeru je $p = 1$, saj smo pridobili le 1 podatek (λ), $k = 5$ in tako je število prostostnih stopenj enako 3. Vrednost testne statistike je, glej [2, str. 2]:

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_{di} - f_{ti})^2}{f_{ti}} = 2.12.$$

Kritična vrednost pri stopnji značilnosti $\alpha = 5\%$ znaša, glej [2, str. 3]:

$$\chi^2_{1-\alpha}(k-p-1) = \chi^2_{0.95}(3) = 7.81.$$

Kot vidimo testna statistika ne preseže kritične vrednosti, kar pomeni, da ničelne domneve H_0 ne moremo zavrniti. Torej ne moremo trditi, da število zadetkov ni porazdeljeno po Poissonu. V praksi torej privzamemo, da je porazdeljeno po Poissonu.
Predpostavimo še, da je število zadetkov posamezne ekipe neodvisno od nasprotne. Verjetnost za posamezni rezultat izračunamo po formuli, glej [1, str. 151]:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y) = \frac{e^{-\lambda_1} \cdot \lambda_1^x}{x!} \cdot \frac{e^{-\lambda_2} \cdot \lambda_2^y}{y!}.$$

United\Liverpool	0	1	2	3	4
0	7.58 %	8.49 %	4.75 %	1.77 %	0.50 %
1	11.06 %	12.39 %	6.94 %	2.59 %	0.73 %
2	8.08 %	9.05 %	5.07 %	1.89 %	0.53 %
3	3.93 %	4.40 %	2.47 %	0.92 %	0.26 %
4	1.43 %	1.61 %	0.90 %	0.34 %	0.09 %

Tabela 4: Verjetnost posameznega rezultata na tekmi Manchester United : Liverpool

V tabeli 4 vidimo verjetnosti posemeznega rezultata na tekmi Manchester United-a in Liverpool-a. Pričakovano je največja verjetnost za rezultat 1:1, saj obe moštvi v povprečju na medsebojni tekmi dosežeta nekaj več kot 1 gol, na splošno pa so rezultati, kjer zmaga United bolj verjetni, saj ima le-ta višjo verjetnost za zadetek.

8 Zaključek

V seminarski nalogi smo si pogledali določene statistične lastnosti angleške Premier lige. Za začetek smo preverili, kako se je razvijalo število zadetkov na tekmi skozi zgodovino in to primerjali z nemško ligo (slika 1) ter za obe izračunali povprečje in standardni odklon. Potem smo si pogledali, kako se v povprečju porazdelijo točke na lestvici (slika 2) in pod drobnogled vzeli prvo mesto, za katerega smo izračunali še interval zaupanja in regresijsko premico. Za tem so sledile domneve, kjer smo dokazali, da lahko z 10% stopnjo tveganja trdimo, če ekipa doseže na tekmi 3 zadetke, v več kot 90% primerih zmaga. Ugotovili smo, da domač teren v resnici predstavlja prednost, saj domače moštvo v 71% tekem ne izgubi. Manjši vpliv na zmago kot domač teren pa ima posest žoge, kjer sicer ekipa z večjo posestjo v 69% res ne izgubi, a primeri, ko je ekipa z mnogo večjo posestjo močno izgubila nam povejo, da posest žoge ni garancija za dober rezultat. Pri zadnji domnevi smo videli vpliv uspešnosti kluba na obiskanost tekem, kjer smo ugotovili, da večjih razlik pri obiskanosti tekem ni. V nadaljevanju smo si pogledali povezanost končnega števila točk ekipe in doseženih zadetkov ter ju povezali z regresijsko premico (slika 5), enako smo storili s streli na gol in številom zadetkov (slika 6). Za konec imamo v tabeli 4 predstavljene še verjetnosti posameznega rezultata na tekmi Manchester United-a in Liverpool-a, glede na njuna prejšnja medsebojna srečanja.

Viri

[1] W. Mendenhall, T. Sincich, Statistics for engineering and the sciences, 5. izdaja, Prentice Hall, 2007

[2] http://courses.wcupa.edu/rbove/Berenson/10th%20ed%20CD-ROM%20topics/section12_5.pdf (Dostopano 16. 2. 2018)

[3] https://www.premierleague.com/stats (Dostopano 10. 1. 2018)

[4] worldfootball.net (Dostopano 10. 1. 2018)
(http://www.worldfootball.net/competition/eng-premier-league/)
(http://www.worldfootball.net/competition/bundesliga/)

[5] http://www2.stetson.edu/~efriedma/research/boldrin.pdf (Dostopano 16. 2. 2018)

[6] http://researchonline.ljmu.ac.uk/3376/3/EvoEPL2_final_send_coauthors_HMS_Last_copy.pdf (Dostopano 16. 2. 2018)