

# Regresijska analiza prihoda nogometnih klubova u Europi

---

**Savatović, Nikola**

**Undergraduate thesis / Završni rad**

**2022**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Economics and Business / Sveučilište u Zagrebu, Ekonomski fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:148:796995>

*Rights / Prava:* [Attribution-NonCommercial-ShareAlike 3.0 Unported/Imenovanje-Nekomercijalno-Dijeli pod istim uvjetima 3.0](#)

*Download date / Datum preuzimanja:* **2024-07-13**



*Repository / Repozitorij:*

[REPEFZG - Digital Repository - Faculty of Economics & Business Zagreb](#)



**Sveučilište u Zagrebu**  
**Ekonomski fakultet**  
**Stručni studij**  
**Računovodstvo i financije**

**REGRESIJSKA ANALIZA PRIHODA NOGOMETNIH  
KLUBOVA U EUROPI**

**Završni rad**

**Ime, prezime i JMBAG studenta : Nikola Savatović, 0081168071**

**Mentor: prof. dr. sc. Mirjana Čižmešija**

**Zagreb, kolovoz, 2022.**

## IZJAVA O AKADEMSKOJ ČESTITOSTI

Izjavljujem i svojim potpisom potvrđujem da je završni rad isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu, a što pokazuju korištene bilješke i bibliografija.

Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog izvora te da nijedan dio rada ne krši bilo čija autorska prava.

Izjavljujem, također, da nijedan dio rada nije iskorišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

---

(vlastoručni potpis studenta)

---

(mjesto i datum)

## SADRŽAJ

1. UVOD	
1.1 Predmet i cilj rada.....	5
1.2. Izvori podataka za empirijsko istraživanje.....	6
2. PRAVILA NATJECANJA I FINANCIRANJA NOGOMETNIH KLUBOVA EUROPE S POSEBNIM OSVRTOM NA ENGLESKU LIGU .....	6
3. OPĆENITO O PREMIJER LIGI.....	8
4. OSNOVNA OBILJEŽJA SEZONE 2020/2021 U ENGLESKOJ LIGI.....	9
5. KLUB SA NAJBOLJOM GOL-RAZLIKOM U ENGLESKOJ LIGI.....	11
6. REGRESIJSKA ANALIZA.....	13
6.1 Regresijski modeli.....	14
6.2. Dijagram rasipanja.....	15
6.3. Model jednostavne linearne regresije.....	16
6.4 Regresijske vrijednosti, rezidualna odstupanja i analiza varijance za model jednostavne linearne regresije.....	21
7. KORELACIJSKA ANALIZA.....	27
8. REZULTATI EMPIRIJSKOG ISTRAŽIVANJA.....	29
8.1 Regresijska i korelacijska analiza gol-razlike i bodova ostvarenih na kraju sezone 2020/2021.....	31
8.2 Regresijska i korelacijska analiza gol-razlike i prihoda klubova na kraju sezone 2020/2021.....	33

9. ZAKLJUČAK.....	34
10. PRILOZI.....	36
11. LITERATURA.....	40

# 1. UVOD

## 1.1. Predmet i cilj rada

Temeljni cilj ovoga rada je uočiti i opisati odnos gol-razlike i pozicije na tablici s prihodima nogometnih klubova diljem Europe da se dokaže da nogometni klubovi koji nisu financijski dominantni, ali imaju dobar rezultat na kraju sezone mogu zaraditi više od klubova bogatih vlasnika sa lošim rezultatom. Također, nastojala se empirijski ispitati pretpostavka da što je bolja gol-razlika bolje je i pozicija na tablici. Regresijskom analizom na realnim podacima ta se pretpostavka i potvrdila. Europa broji čak 55 nogometnih prvih liga, ali svaka je napravljena po istom modelu pošto je svaka od njih pod okriljem krovne europske nogometne organizacije UEFA. U ovome radu istraživat će se Engleska Premier liga primarno jer se nalazi na 1. poziciji po ukupnom broju postignutih bodova klubova u europskim natjecanjima. Također, Engleska liga broji najveći broj svjetski poznatih klubova te je na Otoku uvijek najuzbudljivija završnica sezone zbog najveće izjednačenosti klubova. Engleska Premier liga odraz je Europskog nogometa te začetnik brojnih UEFA-inih ideja i kao takva odabrana je za istraživanje u ovom radu. Najbolji primjer toga jest VAR sistem kojeg je UEFA prvo počela koristiti baš u Engleskoj 2020, a sada se sprovodi u otprilike polovici europskih liga.

U ispitivanju pretpostavke korištena je regresijska i korelacijska analiza i to njen deskriptivno-analitički pristup. Metodama regresijske analize može se numerički izraziti odnos između dviju varijabli te je stoga navedena statistička analiza korištena za kvantificiranje odnosa gol-razlike i broja bodova nogometnih klubova Engleske lige i gol-razlike i prihoda klubova. Korelacijskom analizom prvotno je uočena povezanost između navedenih varijabli, a potom je odabranim numeričkim pokazateljima izmjerena jakost povezanosti. Budući da je uočena pozitivna i jaka povezanost među varijablama, procijenjeni su regresijski modeli koji su povezali prihod kao zavisnu varijablu s gol-razlikom kao nezavisnom varijablom te broj bodova koji ostvaruju klubovi kao zavisnu varijablu i gol-razliku istih klubova kao nezavisnu varijablu. Reprerentativnost regresijskih modela ocijenjena je odabranim mjerama reprerentativnosti. Modelima koji su se pokazali reprerentativnima prognozirane su moguće vrijednosti broja bodova za dane vrijednosti gol-razlike i moguće razine prihoda za pretpostavljene vrijednosti gol-razlike. Provedena je

deskriptivno-statistička analiza modela jednostavne linearne regresije. U analizi je korištena opcija Analiza podataka (engl. *Data Analysis*) u Excel-u.

## 1.2 Izvori podataka

Podaci za empirijsku analizu preuzeti su sa službene sportske statističke stranice Sofascore. Također, dio podataka preuzet je iz časopisa The Sporting News i Interesting football. Velik broj informacija preuzet je sa službene njemačke stranice za vrijednost igrača i klubova Transfermarkt kao i iz par članaka od kojih je najznačajniji ugledni britanski tjednik FourFourTwo. Za teorijski dio o Premier ligi te sezoni 2020/2021 uvelike je pomogla službena stranica Premier lige. Podaci o ukupnim primanjima klubova na kraju sezone preuzeti su iz popularnog Britanskog nogometnog časopisa Mirror.

## 2. PRAVILA NATJECANJA I FINANCIRANJA NOGOMETNIH KLUBOVA EUROPE S POSEBNIM OSVRTOM NA ENGLESKU LIGU

Engleska liga službeno se naziva Premier liga te je po zaradi najprofitabilnija liga na svijetu. Klubovi s dna ljestvice ostvaruju prihode jednake, ako ne i više, od prvaka drugih, svjetski poznatih liga. Vrhunac takve razlike evidentan je bio u sezoni 2013/14 kada je Cardiff, posljednja momčad lige od TV prava ostvario zaradu od 74 milijuna eura što je duplo više od nogometnog velikana Bayern iz Muchena koji je sezonu ranije osvojio i najprestižnije europsko nogometno natjecanje Ligu prvaka, a te sezone osvojio i domaće prvenstvo. Glavni razlog te čudne računice jest popularnost nogometa na „Otoku“ gdje praktički svaki muškarac ima klub za koji strastveno navija za razliku od drugih liga, a poznavajući zakon ponude i potražnje prihod je ekvivalentan popularnosti pa su zbog toga prihodi osrednjih i kvalitetom relativno slabih engleskih klubova tako veliki. Regresijskom analizom odnosa broja bodova i prihoda na kraju sezone potvrđuje se ova povezanost što je veoma logično pošto klub što ima više bodova bolje je rangiran na ljestvici, a što je klub bolje rangiran na ljestvici imati će veći prihod od Premier lige (pri tome se ne misli na ukupni prihod u koji se ubrajaju prihodi od TV prava, prihodi od prodaje dresova i suvenira,

prihodi od ulaznica i sl.) Jedina iznimka je u slučaju jednakog broja bodova kada se promatra gol-razlika. Klub koji ima bolju gol-razliku (broj postignutih pogodaka umanjeno za broj primljenih pogodaka) bit će bolje rangiran na ljestvici. Ukoliko klubovi imaju jednaku gol-razliku bolje rangiran bit će onaj klub koji ima veći broj postignutih pogodaka iz razloga što Engleski nogometni savez više cijeni efikasniju momčad što im na kraju krajeva zbog veće atraktivnosti i gledanosti donosi i više zarade. U slučaju da klubovi imaju jednak broj postignutih i primljenih golova gleda se međusobni omjer što se događa jako rijetko. Međusobni omjer u slučaju kada klubovi imaju jednaku gol-razliku ne gleda se samo u slučajevima kada se radi o klubovima koji se bore za naslov (prvo mjesto) nego i za opstanak u ligi (predzadnje mjesto). Tada klubovi igraju tzv. majstoricu na neutralnom terenu što u povijesti Premier lige još nije viđeno. Najbliže takvom scenariju bili su Manchester United i Newcastle u sezoni 1995/96 kada su čak isprintane ulaznice te rezerviran teren. Na žalost po navijače Newcastlea, United je pitanje pobjednika riješio dva kola prije kraja i tako po deseti put osvojio naslov prvaka Engleske.

Svaki klub bori se za što bolji plasman na kraju sezone prije svega zbog financijskih razloga. Naime Premier liga uređena je tako da velik dio prihoda ovisi o poziciji na ljestvici. Naravno, neki popularni klubovi sa lošijom sezonom i dalje će imati ukupnu zaradu veću od nekog manje popularnog kluba koji je sezonu završio na boljoj poziciji. Ipak, prihod od pozicije igra veliku ulogu te će tako pobjedniku pripasti trideset osam milijuna funti dok će posljednjeplasiranoj momčadi pripasti „tek“ dva milijuna.

Povezanost prihoda s gol-razlikom također je logična s obzirom da se s većom gol-razlikom povećava šansa za boljim plasmanom na ljestvici koja jamči veće prihode. Osim toga, što je bolja gol-razlika kluba, klub će biti popularniji i gledaniji te samim time imati i veće godišnje prihode jer što klub postiže više zгодitaka više ljudi imat će želju potrošiti svoj novac na uživanje u gledanju tih utakmica i golova te kupnju dresa ili nekakvog suvenira iz klupskog dućana. Gol-razlika nije glavni pokazatelj uspješnosti neke momčadi jer ako momčad jednom pobijedi sa 5:0 a idućih pet utakmica izgubi svaku sa minimalnih 1:0 na papiru će se to činiti kao normalna, osrednja gol-razlika dok je u stvarnosti to loš rezultat čak i za klub koji se bori za opstanak u ligi iako gledajući gol-razliku možemo predvidjeti poziciju na ljestvici klubova uz naravno takve netipične situacije. Proučavanjem gol-razlike zadnjih deset godina u Premier ligi opaža se uska



povezanost gol-razlike sa pozicijom na tablici izuzev par iznimaka (npr. klub s vrha ljestvice koji igra „nezanimljiv“ nogomet te postiže malo pogodaka čak i kod najslabijih momčadi te najčešće pobjeđuje sa minimalnom gol-razlikom ili klubovi na dnu ljestvice sa vrlo dobrom gol-razlikom za svoju poziciju koji su imali par odlično odigranih utakmica te pobjeda sa više golova razlike, a kada su gubili gubili su minimalno). Najbolji takav primjer jest Brighton koji je prošlu sezonu završio na niskom 16. mjestu iako je po gol razlici bio 10. najbolji klub Premier lige.

### 3. OPĆENITO O PREMIER LIGI

Premier liga, od strane fanova često nazivana Premiership, profesionana je engleska nogometna liga i najvisi rang u engleskom nogometnom sustavu liga. Peta je najprofitabilija liga u svijetu sportova te najprofitabilnija u Europi sa godišnjom zaradom od 5,3 milijardi dolara. Ispred su samo indijska kriket liga te tri američke lige što je sjajan rezultat uzevši u obzir populaciju Amerike, Indije i Engleske. Po prihodu najbliža europska liga Premier ligi jest Prva njemačka nogometna liga Bundesliga sa godišnjom zaradom od 2,8 milijardi dolara što je skoro upola manje nego prvoplasirana Premier liga. U ligi se natječe dvadeset klubova, a na kraju sezone tri posljednjeplasirane momčadi ispadaju u niži rang Championship dok prve četiri momčadi nastupaju u najjačem europskom natjecanju Ligi prvaka. Glavni sponzor lige je Barclays Banka koja tu uslugu godišnje plaća 350 milijuna britanskih funti pa se liga službeno naziva Barclay's Premier League.

Liga u ovakvom formatu osnovana je 20. veljače 1992. godine kao modernizirana verzija tadašnje Football League koja je osnovana davne 1888. Football League brojala je čak sedamdeset i dva kluba od kojih su tri bila iz Walesa, a ostatak iz Engleske što je značilo preveliki broj utakmica za igrače, a i najboljim i najbogatijim timovima to nije bilo u interesu zbog manjka bitnih utakmica sa kvalitetnim protivnicima koji donose najveću zaradu pa se broj klubova smanjio za čak pedeset. Također, gledanost je bila niska zbog dosta nekvalitetnih utakmica pa je odlučeno da je bolje da najkvalitetniji klubovi igraju češće međusobno nego sa više klubova u sezoni. Liga je uskoro postala najgledanija i najbogatija liga u najpopularnijem sportu na svijetu što zbog strasti Engleza

za ovim sportom, a što zbog spleta okolnosti u kojem je Engleska prva počela igrati nogomet pa samim time ima i najdužu povijest. Također, po uspješnosti u europskim natjecanjima prema UEFA koeficijentima nalazi se na prvom mjestu, ispred španjolske La Lige te talijanske Serie A što im osigurava četiri mjesta u najjačem europskom natjecanju ligi prvaka. Zanimljiv je podatak da pobjednik tog natjecanja dobije 20 milijuna eura dok pobjednik Premier lige osvaja 38 milijuna što je skoro dvostruko više. Od ukupno 45 kluba koja su se natjecala u Premier ligi samo sedam ih je osvajalo naslov od kojih je najuspješniji Manchester United sa čak trinaest naslova. Iako zadnjih godina klubovi iz Premier lige ne postižu značajne rezultate na europskoj sceni, nogometni stručnjaci i dalje englesku ligu smatraju najjačom na svijetu te je opisuju kao ligu koja svake godine ima najveći krug kandidata za naslov što je istina jer ostale lige većinom imaju dva do tri kluba koji konkuriraju za naslov dok u Engleskoj svake sezone najmanje pet klubova konkurira za trofej prvaka što je jedan od glavnih razloga popularnosti i važnosti lige za nogometni svijet.

#### 4. OSNOVNA OBILJEŽJA SEZONE 2020/2021 U ENGLESKOJ LIGI

Za potrebe ovog završnog rada analizirat će se povezanost prihoda sa gol-razlikom te brojem postignutih bodova za posljednju 2020/2021 sezonu. Sezonu je obilježila uvjerljivo najbolja i najefikasnija momčad Manchester City na čelu sa Belgijskim virtuozom Kevinom de Bruyneom koji je s prosječnom ocjenom na kraju svake utakmice 7.78 bio najbolji igrač lige. City je prošle sezone uprihodio 123 milijuna funti od TV prava te četrdeset milijuna funti od lige za osvojeno prvo mjesto. Na začelju, kao i na samom vrhu prošle sezone nije bilo uzbuđenja pošto su tri najlošije momčadi lige Fulham, Sheffield i West Brom bili poprilično daleko od sedamnaestog mjesta koje osigurava opstanak u ligi, te se trenutno bore u Championshipu. Oni su od lige zaradili redom dva, četiri i šest milijuna funti jer Engleski savez ima model po kojem najlošijoj momčadi dodjeljuje dva milijuna funti na kraju sezone dok svakoj sljedećoj bolje rangiranoj daje otprilike dva milijuna više.

Prošle sezone prvi puta u povijesti engleskog nogometa u natjecanju zabilježeno je više pobjeda u gostima nego pobjeda na domaćem terenu. Imali smo prilike vidjeti čak 153 gostujuće pobjede od

kjih je najveća ona Liverpoola protiv Crystal Palacea na Selhurst Park stadionu sa sedam nula te 142 domaće pobjede gdje se ističe pobjeda prvaka Manchester Citya devet nula protiv Southamptona 2. veljače 2021. godine. To je ujedno i pobjeda sa najvećom gol-razlikom u povijesti Premier lige. Najbolju gol-razliku imao je Manchester City sa scoreom +51 dok je najlošiju gol-razliku imao posljednje plasirani Sheffield United koji je primio 42 zgoditka više nego što je zabio što opravdava teoriju da gol-razlika utječe na poziciju na ljestvici.

Što se tiče Hrvatskih nogometaša prošle sezone nastupao je samo dvadesetosmogodišnji Zagrepčanin Mateo Kovačić koji je po statistikama bio šesti najbolji igrač svog Chelsea koji je osvojio solidno treće mjesto zaostavši četrnaest bodova za prvoplasiranim Citijem. Mateo je uspio postići dva pogotka te ostvariti pet asistencija čime je bio treći najbolji asistent momčadi odmah iza Mounta i Jamesa sa po sedam odnosno šest asistencija. Mateo je odradio odličnu sezonu, ali od njega ljudi iz kluba nisu nšta manje niti očekivali obzirom da ga plaćaju pet milijuna funti godišnje.

Najugodnije iznenađenje sezone bio je londonski West Ham završivši na visokom šestom mjestu pod vodstvom iskusnog Davida Moyesa izborivši nastup u 2. najjačem europskom natjecanju Europa ligi iako su po vrijednosti igrača od 350 milijuna na tek jedanaestom mjestu lige. Njihov gradski kolega Arsenal šesti je najvrijedniji klub lige sa ukupnom vrijednošću igrača od 620 milijuna te je od svih klubova najviše razočarao završivši na osmoj poziciji iako su imali plan napasti TOP 4 te tako izboriti nastup u Ligi prvaka. Arsenal je jedna od momčadi koja posljednjih godina često podbacuje te njihovi navijači Gunnersi sigurno žele promjene u klubu no pitanje je vremena kada će se vratiti na stare puteve slave i trofeja.

## 5. KLUB SA NAJBOLJOM GOL-RAZLIKOM U ENGLESKOJ LIGI

Klub sa najboljom gol-razlikom u sezoni 2020/2021 bio je osvajač lige Manchester City što potvrđuje korelaciju između gol-razlike i pozicije na tablici. Klub je to osnovan davne 1800. godine pod imenom St. Marks da bi 1894. promjenio naziv u onaj današnji. Klub je 2008. godine

preuzeo ministar predsjedničkih poslova te zamjenik premijera Ujedinjenih Arapskih Emirata Mansour bin Zayed al Nahyan te u razdoblju nakon toga Građani postižu najbolje rezultate od osnutka kluba osvojivši Premier ligu 4 te FA cup dva puta. Manchester City nastupa na stadionu Etihad od sezone 2003/04. sa kapacitetom od 55 000 mjesta. Stadion nosi ime Etihad zbog sponzorstva sa istoimenom avio kompanijom iz Abu Dhabija što bogati šeici plaćaju pedeset osam milijuna funti svake godine. Klub je u Engleskoj poznat po tome što je jedini klub u povijesti lige koji je završio posljednji pa samim time ispao iz lige sa pozitivnom gol-razlikom što potvrđuje odstupanja u navedenoj regresijskoj analizi. Bilo je to u sezoni 1937/38 kada su postigli 80 golova, a primili 77.

Manchester City klub je sa najvećim prihodima u svijetu nogometa te su prošle godine ostvarili dohodak od 645 milijuna funti. Također, momčad Pepa Guardiole najvrijednija je svjetska momčad sa ukupnom vrijednošću igrača 960 milijuna eura. City ne samo da je imao najbolju gol-razliku već je osvojio titulu najefikasnije momčadi lige sa 83 postignuta pogotka. Također „Građani“ su tokom cijele sezone primili trideset i dva zgoditka što je daleko najmanje od svih momčadi lige što je jedan od glavnih razloga premoćnog osvajanja prvenstva prošle sezone. U trideset osam odigranih kola ubilježili su čak dvadeset sedam pobjeda, pet remija te samo šest poraza što je za konkurentnu ligu poput Premier lige sa punu iznenađenja i neočekivanih rezultata zavidan rezultat. Zanimljiv je podatak da im je najbolji strijelac prošle sezone bio centralni vezni igrač Ilkan Gundogan sa trinaest postignutih pogodaka iz razloga što igrači na njegovoj poziciji rijetko postižu golove. Razlog je taj što trener Pep Guardiola forsira česte rotacije u momčadi i ofenzivnu igru prema naprijed veznjaka. Najbolji asistent u momčadi bio je Belgijac Kevin de Bruyne sa dvanaest uspješnih dodavanja prije zgoditka zauzevši drugo mjesto na ljestvici najboljih asistenata lige odmah iza Britanca Herrya Kane te tako opravdao titulu najplaćenijeg igrača momčadi sa tjednom plaćom od 400 000 funti. Također, de Bruyne je najvrijedniji igrač Citija te prema Transfermarktu, službenoj stranici za vrijednost igrača vrijedi 90 milijuna eura te je tako šesti najvrijedniji nogometaš svijeta.

Ipak, najzaslužniji za osvajanje naslova bio je pedesetjednogodišnji španjolski strateg Jozep Guardiola. Pep je na klupu Citija došao 2016. godine došavši iz Njemačkog velikana Bayerna iz Munchena potpisavši tada rekordan trogodišnji ugovor za trenera sa godišnjom plaćom od petnaest milijuna funti. Pep je tako na listi najplaćenijih svjetskih trenera prestigao Portugalca Josea

Mourinha sa godišnjom plaćom u Manchester Unitedu od četrnaest milijuna funti i postao najplaćeniji svjetski trener. Dulji staž od njega u ligi ima samo Njemac na klupi Liverpoola Jurgen Klop. Španjolac je 2008. te 2010. osvojio titulu najboljeg trenera na svijetu tada vodivši Barcelonu te je tako svojim iskustvom i stručnim znanjem bio ključan za Citijev naslov.

## 6. REGRESIJSKA ANALIZA

Regresijsku analizu čine različite metode ispitivanja ovisnosti jedne zavisne o drugoj nezavisnoj varijabli ili o više drugih nezavisnih varijabli. Varijabla čije se varijacije objašnjavaju pomoću drugih varijabli naziva se zavisnom varijablom ( $Y$ ), a varijable kojima se objašnjava varijacija zavisne varijable nazivaju se nezavisnim varijablama ( $X_j, j = 1, 2, 3 \dots k$ ).

Regresijska analiza je algoritam prediktivnog modeliranja koji s velikom sigurnošću predviđa rezultat varijable i identificira (nezavisne) varijable koje utječu na varijablu ishoda (zavisna varijabla). Sažeto rečeno, to je način pronalaženja odnosa između nezavisnih varijabli i zavisne varijable za dobivanje regresijskog modela. Različita su područja primjene regresijske analize, ponajprije za kvantificiranje odnosa između varijabli odnosno procjenom regresijskog modela, a potom i za prognoziranje vrijednosti zavisne varijable za dane vrijednosti nezavisne/nezavisnih varijabli ako je procijenjeni model reprezentativan. Regresijska analiza korisna je kada npr. treba predvidjeti plaću zaposlenika ili prihod tvrtke za određeni broj zaposlenih. Također, regresijska analiza može predvidjeti stopu rasta kupca u bilo kojoj industriji ili primjerice hoće li uvođenje novog proizvoda na tržište biti uspješno ili ne. Moderne metode regresijske analize mogu predvidjeti hoće li neki kupac kupiti proizvod ili ne te različite oblike prijevara i neželjene pošte.

Postoje mnoge vrste regresijskih tehnika koje se koriste uzimajući u obzir različite čimbenike i ishode. Ovisno o obliku veze između varijabli, najčešće se koriste linearna, logistička te polinomna regresija. Linearna regresija koristi se kada je zavisna varijabla linearno povezana s nezavisnim varijablama (u parametrima). Od svih regresijskih oblika ona se najranije krenula detaljno proučavati i ekstenzivno koristiti u praktičnim primjenama. Glavni razlog tomu jest jednostavnost modeliranja odnosa između varijabli.

Regresijska analiza može se provoditi kao deskriptivna i kao inferencijalna. Ako se analizirani podaci tretiraju kao populacija (svi mogući podaci) tada se provodi deskriptivno statistička analiza. Zaključci se odnose isključivo na analizirani skup podataka. U inferencijalnom pristupu analizirani podaci su uzorak, a zaključci se (uz ispunjenje određenih pretpostavki) poopćuju na cijelu populaciju.

Za potrebe završnog rada procijeniti će se parametri za dva modela jednostavne linearne regresije. U prvom modelu zavisna varijabla je broj bodova, a nezavisna varijabla je gol-razlika klubova na kraju sezone. U drugom modelu zavisna varijabla je prihod klubova u prvoj engleskoj nogometnoj ligi na kraju sezone 2020/2021 za ostvarenu poziciju na tablici, a gol-razlika klubova je nezavisna varijabla.

## 6.1. Regresijski modeli

Regresijski model je osnova svake regresijske analize. To je algebarski model kojim se analitički izražava statistički odnos među pojavama. Oni se primjenjuju u gotovo svim stručnim i znanstvenim područjima.

Model može povezivati dvije varijable tako da bude *deterministički (funkcionalan)* ili tako da bude *statistički (stohaistički)*.

*Determinističkim modelom* opisuje se egzaktna veza među varijablama (za svaku vrijednost nezavisne varijable  $X$  jednoznačno je određena vrijednost zavisne varijable  $Y$ ). To je model u kojem je svaki skup varijabilnih stanja jedinstveno određen parametrima u modelu i skupovima prethodnih stanja tih varijabli bez neizvjesnosti i neočekivanih događaja. Primjer determinističkog modela u nogometnom svijetu bio bi broj pobjeda i broj bodova gdje pobjeda uvijek nosi tri boda gdje bi  $Y$  označavao broj pobjeda, a  $X$  broj ostvarenih bodova. Opaža se da bi  $X$  varijabla uvijek bila tri puta veća od vrijednosti varijable  $Y$ .

Funkcionalna (deterministička) veza je oblika:

$$Y = f(X) \quad (1)$$

*Statistički odnosi* pojava pod utjecajem su nesistematskih, stohastičkih varijacija, čija prisutnost izvire iz prirode tih odnosa. Svaki regresijski model sadrži slučajnu varijablu koja sadrži nesistematske utjecaje na zavisnu varijablu i po kojoj se statistički model razlikuje od determinističkog modela. Jednosatavnije rečeno, statistički model opisuje situaciju u kojoj postoji neizvjesnost te ga karakterizira neki stupanj slučajnosti. Primjer statističkog modela je upravo model koji će se analizirati u radu. Zavisna varijabla je prihod koji je određen gol-razlikom ali i slučajnom varijablom u kojoj su sadržani svi ostali faktori koji neizravno utječu na prihod, a nisu sadržani u niti jednoj drugoj nezavisnoj varijabli.

*Statistički model* izražava slabiju vezu između varijabli.

Vrijednost zavisne varijable  $Y$  nije jednoznačno određena za zadanu vrijednost nezavisne varijable  $X$ . Statistička veza je oblika:

$$Y = f(X) + u \quad (2)$$

pri čemu je  $f(x)$  – funkcionalna (deterministička) komponenta

$u$  – slučajna varijabla

$Y$  – zavisna varijabla

$X$  nezavisna varijabla tj. pojava čiji se utjecaj na pojavu  $Y$  ispituje

## 6.2. Dijagram rasipanja

Prvi je korak u regresijskoj analizi dviju varijabli konstruiranje dijagrama rasipanja. Pomoću njega se zorno uočava priroda odnosa dvije varijable predočenim empirijskim vrijednostima dviju varijabli. Koristi se kako bi se na jednostavan vizualan način ponajprije uočilo postoji li povezanost između dvije varijable. Specifičan raspored točaka na dijagramu rasipanja daje jasne indikacije o kakvoj vrsti povezanosti se radi (ako ona uopće postoji). Dijagram se crta tako da se u koordinatni sustav unose parovi vrijednosti varijable  $X$  i  $Y$  tj. sastoji se od  $n$  točaka  $(x_i, y_i)$ . Iz rasporeda točaka u dijagramu rasipanja zaključuje se o obliku, smjeru i jakosti veze.

Raspoređuju li se točke u dijagramu rasipanja od donjeg lijevog kuta kvadranta prema gornjem desnom kutu, odnos između pojava može se analitički izraziti pomoću modela jednostavne linearne regresije. Što je rasipanje oko pravca manje tj. što su unesene točke bliže zamišljenom pravcu to je povezanost između varijabli veća. Čest je i krivolinijski model regresije u kojem se točke u dijagramu rasipanja raspoređuju oko neke krivulje. Raspoređivanje točaka može biti takvo da govori o nepostojanju povezanosti među varijablama. U tom slučaju je procjenjivanje parametara u regresijskom modelu nesmisleno.

Uz pomoć dijagrama rasipanja donosi se odluka o mogućoj povezanosti analiziranih varijabli. Ako se ustanovilo da povezanost postoji, određuje se oblik (najčešće linearni), a zatim smjer veze odnosno je li linearna veza pozitivna ili negativna. Veza je pozitivna ako porast vrijednosti nezavisne varijable prati linearni ili približno linearni porast vrijednosti zavisne varijable. U slučaju kada se poveća vrijednost nezavisne varijable, a vrijednost se zavisne varijable linearno ili približno linearno smanjuje, riječ je o negativnom smjeru linearne veze.

Dijagram rasipanja prvi je otkrio i njime se koristio engleski znanstvenik John Frederick W. Herschel 1833.godine proučavavši orbitu zvijezda. Točnije, usporedio je kut zvijezda u orbiti sa godinom mjerenja te tako otkrio povezanost. Ipak, za porast popularnosti dijagrama rasipanja u daljoj povijesti najzaslužniji je engleski statističar Francis Galton, prozvan ocem moderne statistike. On je jedan od tvoraca statističkog koncepta korelacije te autor preko tristo radova i knjiga.



### 6.3. Model jednostavne linearne regresije

Metodama regresijske analize moguće je opisati odnos između varijabli  $X$  i  $Y$ . Ako točke prikazane dijagramom raspršenosti izgledaju kao da su raspoređene oko i uz neki pravac, onda se radi o jednostavnoj linearnoj regresiji. Model jednostavne linearne regresije prikladan je za opisivanje odnosa između pojava, te je za nj karakteristično da promjenu jedne pojave prati približna ako ne i identična linearna promjena druge. Za potrebe ovoga završnog rada analizirat će se odnos gol-razlike i bodova na kraju sezone kao i ukupni prihodi od Saveza. Pođe li se od pretpostavke da je riječ o linearnom odnosu između dvije varijable, funkcija  $f(x)$  bit će sljedećeg oblika:

$$f(x) = a + bx \quad (3)$$

Uzimajući u obzir opći oblik modela i oblik linearne funkcije, model jednostavne linearne regresije postaje:

$$Y = a + bX + u \quad (4)$$

U navedenom modelu  $X$  je nezavisna, a  $Y$  zavisna varijabla. Varijabla  $u$  izražava nepoznate i apstrahirane utjecaje na varijaciju varijable  $Y$  i predstavlja odstupanje od funkcionalnog odnosa. Zbog toga se često naziva i greškom relacije. Njezina prisutnost u modelu je posljedica postojanja statističke povezanosti pojava.

U modelu (4)  $a$  i  $b$  su parametri. Za procjenu parametara u modelu potrebno je raspolagati s empirijskim podacima odnosno s  $n$  parova vrijednosti zavisne i nezavisne varijable. Kada bi odnos između varijabli bio funkcionalan, u modelu ne bi bila prisutna varijabla  $u$ . Ako se pretpostavi linearni odnos između varijabli parovi kojih su vrijednosti točke u dijagramu rasipanja, onda se između tih točaka može smjestiti beskonačno mnogo pravaca. Izbor pravca ili veličine parametara  $a$  i  $b$ , odnosno funkcije  $f(x)$ , mora početi od određenog kriterija izbora. Ima više kriterija koji služe toj svrsi odnosno postoji više različitih metoda procjene parametara u modelu jednostavne linearne regresije. Izbor se može provesti pođe li se od nastojanja da odstupanja stvarnih vrijednosti zavisne varijable od vrijednosti funkcije budu mala. Odstupanja se mogu mjeriti na različite načine. Uobičajeno se mjere odstupanja usporedna s osi ordinate (s vertikalnom osi).

Analiza modela (4) oslanja se na stvarne vrijednosti varijabli. Polazni model jednostavne linearne regresije za skup od  $n$  vrijednosti  $(x_i, y_i)$  varijabli  $X$  i  $Y$  može se napisati i kao sustav jednadžbi :

$$y_i = a + bx_i + u_i \quad i = 1, 2, 3 \dots n, \quad (5)$$

ili alternativno :

$$\begin{aligned} y_i &= \hat{y}_i + u_i \\ \hat{y}_i &= a + bx_i \end{aligned} \quad (6)$$

gdje je  $\hat{y} = a + bx$  procijenjena odnosno regresijska vrijednost zavisne varijable za danu vrijednost nezavisne varijable. Iz izraza (6) odstupanje  $u_i$  je razlika stvarne (empirijske

vrijednosti zavisne varijable) i po modelu procijenjene (regresijske) vrijednosti zavisne varijable i zapisuje se:

$$u_i = y_i - \hat{y}_i \quad (7)$$

odnosno:

$$u_i = y_i - a - bx_i \quad (8)$$

Veličine  $u_i$  u prethodnom izrazu nazivaju se rezidualnim odstupanjima. Ima ih toliko koliko i opažanja ( $n$ ). Dobar primjer je momčad Man Citija koja je u sezoni 1937/38 imala pozitivnu gol-razliku i po tom pitanju bila bolja od pola lige što im na kraju nije pomoglo kod plasmana gdje su završili posljednji. To je jedini takav ekstreman primjer u povijesti Prve Engleske nogometne lige, ali potvrđuje da barem kod regresijske analize gol-razlike i pozicije na tablici nema striktnog pravila i moguća su odstupanja i nepravilnosti.

Iz (6) je vidljivo da je stvarna vrijednost zavisne varijable zbroj vrijednosti linearne funkcije i odstupanja  $u$ . S obzirom da je tih odstupanja koliko i parova vrijednosti, izbor funkcije odnosno pravca mogao bi se temeljiti na veličini njihovog zbroja. Taj zbroj nije dobra osnovica za izbor adekvatne linearne funkcije jer se moguća velika odstupanja s pozitivnim predznakom mogu potirati s velikim odstupanjima negativnog predznaka pa će zbroj odstupanja biti mali iako je riječ o pravcu slabe reprezentativnosti. Umjesto zbroja moguća osnovica izbora je zbroj kvadrata rezidualnih odstupanja, odnosno zbroj kvadrata odstupanja stvarnih vrijednosti zavisne varijable od regresijskih vrijednosti linearne funkcije  $\hat{Y}$ . Za parametre te funkcije mogu se odrediti veličine za koje rezidualni zbroj kvadrata doseže minimum. Metoda određivanja parametara je prvi zadatak u statističkoj analizi regresijskog modela. Izabere li se metoda najmanjih kvadrata tada je polazni izraz rezidualni zbroj kvadrata:

$$SQ = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (a + bx_i))^2 \quad (9)$$

Kako su u navedenom izrazu veličine  $y_i$  i  $x_i$  dane, rezidualni zbroj kvadrata ovisi o parametrima  $a$  i  $b$ . Da bi došlo do ocjena parametara  $a$  i  $b$  u smislu metode najmanjih kvadrata, nužno je primijeniti postupak određivanja minimuma funkcije  $SQ$ . Iz prvog uvjeta minimuma dolazi do sljedećeg sustava jednadžbi :

$$\frac{\delta SQ}{\delta a} = -2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \quad (10)$$

$$\frac{\delta SQ}{\delta b} - 2 \sum_{i=1}^n (y_i - a - bx_i)x_i = 0 \quad (11)$$

Uredi li se sustav postaje:

$$na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (12)$$

$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \quad (13)$$

Jednadžbe (12) i (13) zovu se normalnim jednadžbama. Normalne jednadžbe linearne su u parametrima, a sadrže dvije nepoznaice, koliko je i nepoznatih parametara. Rješenje sustava normalnih jednadžbi je:

$$b = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \quad (14)$$

$$a = \bar{y} - b \bar{x} \quad (15)$$

Model jedostavne linearne regresije s procijenjenim parametrima ima oblik:

$$\hat{y} = a + bx \quad (16)$$

U ovom modelu s procijenjenim parametrima,  $a$  je konstanti član. To je vrijednost regresijske funkcije za vrijednost zavisne varijable  $x = 0$ .  $b$  je regresijski koeficijent. On pokazuje za koliko se linearno mijenja vrijednost regresijske funkcije za jedinični porast vrijednosti nezavisne varijable  $X$ . Predznak koeficijenta je pozitivan ili negativan. Ako je parametar  $b$  pozitivan, znači da će s jediničnim porastom nezavisne varijable, zavisna varijabla linearno i u prosjeku porasti za  $b$ . Ako je  $b$  negativan, s jediničnim porastom nezavisne varijable, zavisna varijabla će se linearno i u prosjeku smanjiti za  $b$ .

#### 6.4. Regresijske vrijednosti, rezidualna odstupanja i analiza varijance za model jednostavne linearne regresije

Pomoću regresijske jednadžbe s ocijenjenim parametrima utvrđuju se regresijske vrijednosti. Regresijske su vrijednosti dane izrazom :

$$\hat{y}_i = a + bx_i \quad (17)$$

Izračunavaju se tako da se uvrste empirijske vrijednosti nezavisne varijable  $X$  u jednadžbu s ocijenjenim parametrima. Te vrijednosti predstavljaju ocjenu razine zavisne varijable za dane stvarne vrijednosti nezavisne varijable. S obzirom na to da se analizira statistički odnos između pojava, stvarne vrijednosti zavisne varijable  $Y$  razlikuju se od regresijskih vrijednosti. Kako je već rečeno, ta razlika izračunata na osnovi modela s ocijenjenim parametrima predstavlja rezidualno odstupanje. Rezidualno odstupanje jest ocjena grešaka relacije u polaznom modelu jednostavne linearne regresije. Rezidualno odstupanje može poprimiti pozitivne ili negativne vrijednosti, a može biti i jednako nuli. Ako je rezidualno odstupanje pozitivno, znači da je stvarna vrijednost procijenjenim regresijskim modelom podcijenjena i obrnuto.

Jednadžba jednostavne linearne regresije s ocijenjenim parametrima metodom najmanjih kvadrata aproksimira odnos između varijabli u smislu aritmetičke sredine. Zbog toga vrijede i sljedeća svojstva ocijenjenog modela:

- a) Zbroj odstupanja stvarnih vrijednosti zavisne varijable  $y$  od regresijskih vrijednosti  $\hat{y}$  jednak je 0,
- b) Zbroj kvadrata tih odstupanja je minimalan i
- c) Zbroj produkata regresijskih vrijednosti i rezidualnih odstupanja jednak je 0. Zbroj produkata vrijednosti nezavisne varijable i rezidualnih odstupanja također je jednak 0.

Kako je zbroj odstupanja empirijskih vrijednosti zavisne varijable od njezinih regresijskih vrijednosti jednak 0, to je aritmetička sredina stvarnih vrijednosti te varijable jednaka aritmetičkoj sredini regresijskih vrijednosti.

Rezidualna odstupanja dana su izrazom:

$$u_i = y_i - \hat{y}_i \quad (18)$$

Vrijednosti rezidualnih odstupanja izražene su u mjernim jedinicama zavisne varijable. Radi jednostavnije prosudbe njihovih obilježja računaju se i relativna i standardizirana rezidualna odstupanja.

Relativna rezidualna odstupanja dobivaju se dijeljenjem tih odstupanja pripadajućom stvarnom vrijednošću zavisne varijable, a omjer se potom pomnoži sa 100. Izraz za relativno izražena rezidualna odstupanja je:

$$u_{iREL} = \frac{y_i - \hat{y}_i}{y_i} 100 \quad (19)$$

Standardizirana odstupanja računaju se dijeljenjem rezidualnih odstupanja regresijskom standardnom devijacijom odnosno standardnom greškom.

U pravilu, ocijenjeni model je reprezentativniji što su manja rezidualna odstupanja. Rezidualnih odstupanja ima koliko i regresijskih vrijednosti, odnosno parova vrijednosti varijabli  $X$  i  $Y$ . Ona predstavljaju disperziju u odnosu na regresijsku funkciju kao prosječnu veličinu.

Stupanj varijacije stvarnih vrijednosti zavisne varijable u odnosu na procijenjene vrijednosti pomoću regresije mjeri se različitim mjerama od kojih je najvažnija varijanca regresije i iz nje

izvedena standardna devijacija regresije te koeficijent varijacije regresije. Mjere disperzije oko regresije služe između ostalih pokazatelja za ocjenu kvalitete odnosno reprezentativnosti modela.

Statističko – analitičke veličine za prosudbu reprezentativnosti regresije temelje se na raščlanjivanju zbroja kvadrata odstupanja vrijednosti zavisne varijable od njezina prosjeka, kako je dano jednadžbom (20).

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (20)$$

Jednadžba (20) temelj je analize varijance. Ta jednadžba predstavlja razdiobu zbroja kvadrata odstupanja empirijskih vrijednosti zavisne varijable od njezine aritmetičke sredine ( ukupni zbroj kvadrata). Prva komponenta, odnosno prvi član s desne strane jednadžbe je zbroj kvadrata odstupanja regresijskih vrijednosti od njihove aritmetičke sredine. Ta se komponenta naziva i sumom kvadrata odstupanja protumačenih modelom. Drugu komponentu čini zbroj kvadrata odstupanja empirijskih vrijednosti zavisne varijable od regresijskih vrijednosti. Ta se komponenta alternativno naziva neprotumačenom sumom kvadrata tj. rezidualnom sumom kvadrata.

Ako se svaki član u jednadžbi (20) podijeli ukupnim zbrojem kvadrata i ako se strane jednadžbe zamijene dolazi do raščlambe ukupnog zbroja kvadrata u relativnom iznosu :

$$\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} + \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2} = 1 \quad (21)$$

Prvi član s lijeve strane navedene jednadžbe naziva se koeficijentom determinacije, a drugi član je koeficijent alijenacije.

U razvijenom obliku zbrojevi kvadrata dani su jednadžbama :



$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2 \quad (22)$$

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = a \sum_{i=1}^n y_i + b \sum_{i=1}^n x_i y_i - n\bar{y}^2 \quad (23)$$

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n y_i^2 - a \sum_{i=1}^n y_i - b \sum_{i=1}^n x_i y_i \quad (24)$$

Polazna veličina za mjerenje disperzije je rezidualni zbroj kvadrata. Varijanca regresije je prosječni rezidualni zbroj kvadrata.

$$\sigma_{\hat{y}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (25)$$

Standardna devijacija regresije je drugi korijen iz varijance regresije:

$$\sigma_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (26)$$

Standardna devijacija regresije pokazuje koliko je prosječno odstupanje empirijskih vrijednosti zavisne varijable od njezinih regresijskih vrijednosti. Izražena je u istim mjernim jedinicama kao i zavisna varijabla, pa je po tome apsolutna mjera disperzije oko regresije. Relativna mjera disperzije oko regresije je koeficijent varijacije regresije. Koeficijent varijacije je omjer standardne devijacije regresije i aritmetičke sredine zavisne varijable pomnožen sa sto, što se zapisuje:

$$V_{\hat{y}} = \frac{\sigma_{\hat{y}}}{\bar{y}} \cdot 100 \quad (27)$$

Varijanca, tj, standardna devijacija i koeficijent varijacije ovise o rezidualnom zbroju kvadrata. Kada bi svako rezidualno odstupanje bilo jednako nuli i rezidualni zbroj kvadrata bio bi jednak nuli i u tom slučaju vrijednosti regresije bile jednake vrijednostima linearne funkcije. Odnos između varijabli u tom primjeru nebi bio statistički nego deterministički. U načelu, mala vrijednost tih pokazatelja može biti pokazatelj dobre reprezentativnosti analitičkog izraza odnosa tih pojava i obrnuto.

Pokazatelj reprezentativnosti regresije je i koeficijent determinacije. Koeficijent determinacije je omjer protumačenog i ukupnog zbroja kvadrata. Dan je jednadžbom:

$$r^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad 0 \leq r^2 \leq 1 \quad (28)$$

alternativno:

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (29)$$

Koeficijent determinacije varira u zatvorenom intervalu od nule do jedan. Jednak je nuli ako je protumačen zbroj kvadrata jednak nuli. Poželjna je velika vrijednost tog koeficijenta (blizu jedan), jer to znači da je mala vrijednost rezidualnog zbroja kvadrata, a time je mala i disperzija oko regresije. Kao analitički pokazatelj u prosudbi kakvoće regresije služi pored koeficijenta determinacije i korigirani koeficijent determinacije. Korigirani koeficijent determinacije dan je izrazom:

$$\bar{r}^2 = 1 - \frac{n-1}{n-2}(1-r^2) \quad (30)$$

Korigirani koeficijent determinacije jednak je ili manji od koeficijenta determinacije.

## 7. KORELACIJSKA ANALIZA

Korelacijska analiza sastoji se u primjeni metoda kojima se utvrđuju pokazatelji jakosti ili jakosti i smjera statističkih veza među pojavama. Najjača je funkcionalna povezanost pojava. Ako je statistički odnos pojava oblikom linearan, razlikovat će se i smjer veze, koji može biti pozitivan i negativan.

Ako porast jedne pojave prati porast druge pojave ili ako pad jedne pojave prati pad druge pojave, govori se o pozitivnoj korelaciji.

Ako porast jedne pojave prati pad druge pojave i obrnuto riječ je o negativnoj korelaciji.

Kada su podaci o pojavama dani kao oblici redosljednog obilježja, njihovu povezanost moguće je mjeriti pokazateljem korelacije ranga.

Ako su pojave predočene oblicima nominalnih obilježja, za analizu ovisnosti upotrijebit će se mjere asocijacije.

Osnovni zadatak analize stupnja jakosti statističke veze među pojavama sastoji se od utvrđivanja prikladnih pokazatelja, odnosno koeficijenta korelacije i asocijacije. U analizi smjera i jakosti povezanosti između dviju numeričkih varijabli uobičajeno se koristi Pearsonov koeficijent korelacije.

Pearsonov koeficijent linearne korelacije je numerički pokazatelj smjera i jakosti veze između dviju numeričkih varijabli i izračunava se po izrazu:

$$r = b \frac{\sigma_x}{\sigma_y} \quad (31)$$

Ako je prethodno izračunan koeficijent determinacije, koeficijent linearne korelacije je drugi korjen iz koeficijenta determinacije. Predznak koeficijentu korelacije pridružuje se ovisno o predznaku parametra  $b$  u procijenjenom regresijskom modelu. Ako je  $b$  pozitivan, koeficijent korelacije je pozitivan i povezanost između varijabli je pozitivna i obrnuto.

Kako je već rečeno, koeficijentom linearne korelacije mjeri se jakost i smjer statističke povezanosti dviju pojava predočenih u obliku parova vrijednosti numeričkih varijabli. Koeficijent varira u zatvorenom intervalu od minus jedan do plus jedan. Prosuđivanje jakosti veze pomoću koeficijenata linearne korelacije kao deskriptivno – statističke veličine valja povezivati s veličinom koeficijenta determinacije. Orijentacijski u tome mogu poslužiti sljedeći odnosi navedeni u tablici 2.

Tablica 1. Odnos koeficijenta determinacije i koeficijenta korelacije

Koeficijent determinacije	Apsolutna vrijednost koeficijenta linearne korelacije	Tumačenje
$r^2$	$ r $	
0	0	Odsutnost korelacije
0.00 – 0.25	0.00 – 0.50	Slaba korelacija
0.25 – 0.64	0.5 – 0.8	Korelacije srednje jačine
0.64 – 1	0.8 – 1	Čvrste korelacije
1	1	Potpuna (perfektna) korelacija

Izvor: izrada autora prema Čižmešija, M., Kurnoga Živadinović, N. (2015). Uvod u gospodarsku statistiku (priručnik s pregledom formula, 3. izdanje). Element, Zagreb

## 8. REZULTATI EMPIRIJSKOG ISTRAŽIVANJA

U prvom dijelu empirijskog istraživanja analizirat će se odnos gol-razlike i bodova na kraju sezone dok je drugi dio posvećen analizi odnosa gol-razlike i prihoda klubova od nacionalnog saveza za ostvarenu poziciju na tablici.

Do podataka za gol-razliku klubova i broja bodova došlo se uz pomoć Sofascore sportske aplikacije, dok se do informacija o prihodima klubova došlo zahvaljući objavljenom članku poznatog engleskog časopisa Mirror iako uzevši u obzir da je nogomet najpopularniji sport na svijetu, a Engleska liga najgledanija do podataka se moglo doći i na mnogo drugih načina.

Raspolaže se podacima iz sezone 2020/21 u kojoj je odigrano trideset osam kola. U svakom kolu odigrano je po deset utakmica što je ukupno tristo osamdeset utakmica u cijeloj sezoni. Smatra se da je to relevantan broj utakmica za reprezentativnu regresijsku analizu.

Regresijska analiza provedena je za dvadeset klubova što znači da se rasplaže s dvadeset parova vrijednosti zavisne i nezavisne varijable. Podaci za analizu dani su u tablici 2.

Tablica 2. Ime kluba, pozicija na tablici, gol-razlika, ostvareni broj bodova klubova i ostvareni prihod u sezoni 2020/2021.

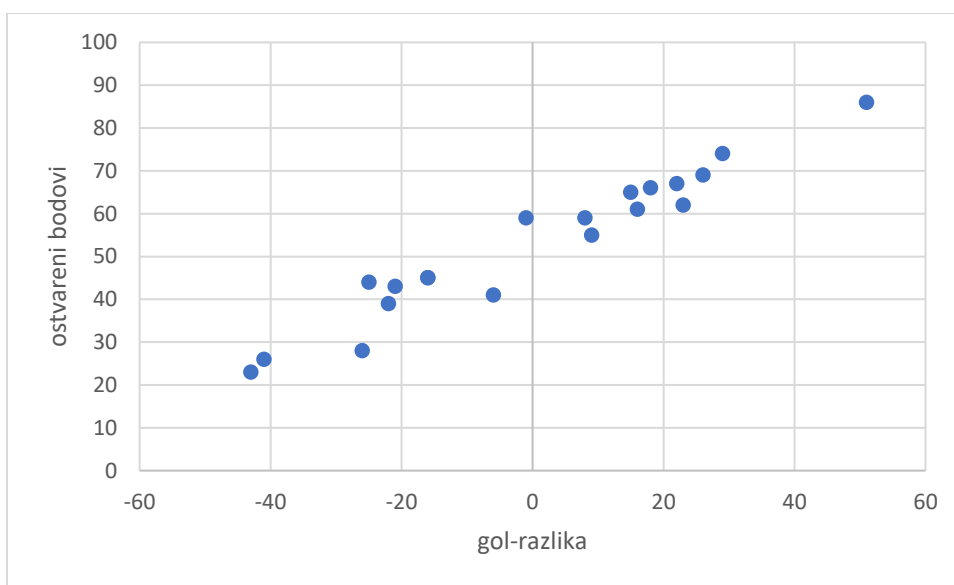
/Ime nogometnog kluba	Pozicija na tablici	Gol-razlika	Ostvareni bodovi na kraju sezone	Ostvareni prihod u milijunima funti
Manchester City	1.	51	86	38
Manchester United	2.	29	74	36
Liverpool	3.	26	69	34
Chelsea	4.	22	67	32
Leicester City	5.	18	66	30
West Ham United	6.	15	65	28
Tottenham Hotspur	7.	23	62	27
Arsenal	8.	16	61	24
Leeds United	9.	8	59	23
Everton	10.	-1	59	21
Aston Vila	11.	9	55	19
Newcastle United	12.	-16	45	17
Wolverhampton	13.	-16	45	15
Crystal Palace	14.	-25	44	13
Southampton	15.	-21	43	11
Brighton	16.	-6	41	10
Burnley	17.	-22	39	8.5
Fulham	18.	-26	28	6
West Bromwich Albion	19.	-41	26	4
Sheffield United	20.	-43	23	2.5

Izvor: [interestingfootball.com/premier-league-prize-money-epl-teams-prize-by-position-table/](https://interestingfootball.com/premier-league-prize-money-epl-teams-prize-by-position-table/)

### 8.1. Regresijska i korelacijska analiza gol-razlike i bodova ostvarenih na kraju sezone 2020/2021.

Bodovi ostvareni na kraju sezone o ovom primjeru su zavisna varijabla dok je postignuta gol-razlika nezavisna varijabla. Pretpostavlja se da o njoj ovisi koliko će koji klub skupiti bodova.

Grafikon 1. Dijagram rasipanja za gol-razliku i bodove koje su klubovi ukupno ostvarili na kraju sezone 2020/2021



Izvor: izrada autora

Proučavanjem dijagrama rasipanja konstatira se da između gol-razlike i ostvarenih bodova na kraju sezone postoji očigledna i pozitivna linearna i jaka povezanost.

Regresijska analiza provedena je u programu Excel uporabom opcije Analiza podataka (engl. *Data analyses*). Rezultati obrade dani su u prilogu.

Model jednostavne linearne regresije s procijenjenim parametrima je:

$$\hat{y}_i = 52,85 + 0,64x_i$$

Iz navedenoga slijedi da će se za svako povećanje gol-razlike za jedan, broj bodova kluba linearno i u prosjeku povećati za 0,64 što sažetije rečeno znači da će svaki postignuti pogodak klubu u prosjeku donjeti 0.64 boda gdje također svaki primljeni zgoditak znači gubitak u prosjeku 0,64 boda. Ovdje treba spomenuti kako je nemoguće ostvariti 0,64 odnosno izgubiti isto toliko bodova pošto postoje varijante samo od tri boda za pobjedu, jedan bod za remi te nula bodova za poraz, ali je svejedno itekako reprezentativan i bitan podatak kako nogometnim statističarima tako i trenerima jer im olakšava koliko značenje ima svaka pobjeda, remi i poraz. Osim toga, mjere reprezentativnosti koje su dane u nastavku pokazuju da je procjenjeni regresijski model reprezentativan što znači da se osim opisivanja odnosa između analiziranih varijabli on može koristiti i u prognostičke svrhe.

Prema procijenjenom modelu slijedi da ako je gol-razlika nekog kluba npr. devetnaest, očekuje se da će taj klub na kraju sezone ostvariti 65,01 bodova što je usporedivši tu gol-razliku sa gol-razlikom petoplasiranog Leicestera od osamnaest i više nego reprezentativno uzevši u obzir podatak da su oni s tom gol-razlikom ostvarili 66.

$$\hat{y}_l = 52,85 + 0,64 \cdot 19 = 65,01$$

Ako recimo neki klub sezonu završi sa vrlo negativnom gol-razlikom od -40 nakon uvrštavanja u model dobiva se podatak da se za taj klub očekuje 27,25 ostvarenih bodova na kraju sezone što je veoma blizu West Bromu koji je sezonu završio sa gol-razlikom od -41 te 26 ostvarenih bodova.

$$\hat{y}_l = 52,85 + 0,64 \cdot (-40) = 27,25$$

Prema dobivenim rezultatima koeficijent linearne korelacije iznosi 0,9696 što znači da između gol-razlike i ostvarenih bodova postoji pozitivna i vrlo jaka povezanost.

Procjenjenim modelom protumačeno je 94% (koeficijent determinacije je 0,94) svih odstupanja što znači da je regresijski model reprezentativan i njegovo korištenje u prognostičke svrhe smatra se opravdanim.



Osim toga, standardna devijacija regresije izračunana po izrazu (26) iznosi 4,03 bodova

$$\sigma_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} = \sqrt{\frac{324,795}{20}} = 4,03$$

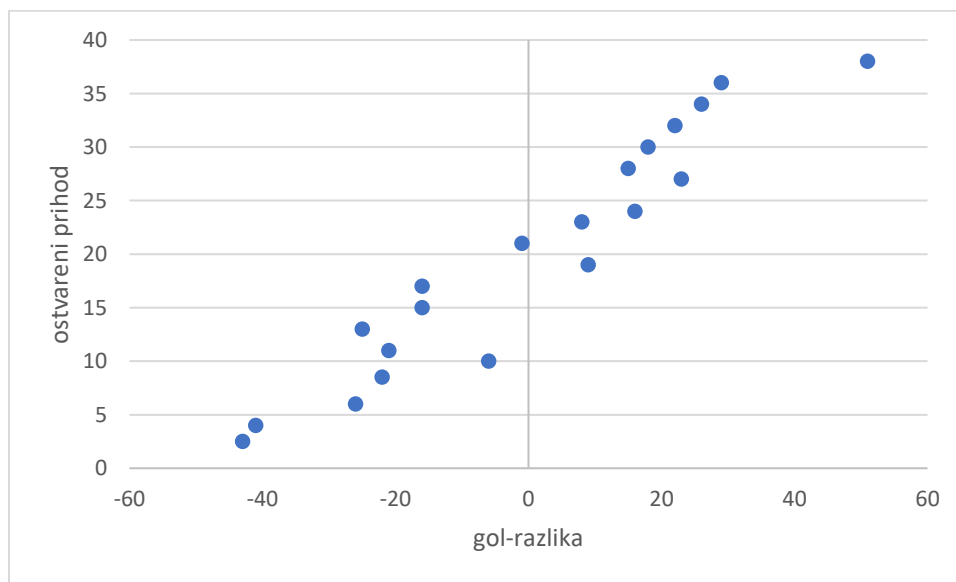
Koeficijent varijacije regresije izračunan po izrazu (27) je 20,20%

$$V_{\hat{y}} = \frac{\sigma_{\hat{y}}}{\bar{y}} \cdot 100 = \frac{4,03}{52,85} \cdot 100 = 20,20\%$$

Zaključuje se da je prosječno odstupanje stvarnih od regresijskih (po modelu procijenjenih vrijednosti) broja bodova 4,03 boda ili 20,20%. Zaključuje se da je model i po ovim pokazateljima reprezentativnosti zadovoljavajuće reprezentativan.

## 8.2. Regresijska i korelacijska analiza gol-razlike i prihoda klubova na kraju sezone 2020/2021

Grafikon 2. Dijagram rasipanja za gol-razliku i prihoda koje su klubovi ostvarili od Engleskog nogometnog saveza na kraju sezone 2020/2021



Izvor: izrada autora

Prihod ostvaren na kraju sezone u ovom slučaju je zavisna varijabla dok je ostvarena gol-razlika nezavisna. Pretpostavlja se da gol-razlika, između ostaloga, određuje visinu prihoda klubova.

Temeljem dijagrama rasipanja zaključuje se da između gol-razlike i prihoda na kraju sezone postoji pozitivna linearna i jaka povezanost.

Model jednostavne linearne regresije s procijenjenim parametrima je

$$\hat{y}_i = 19,95 + 0,41x_i$$

Zaključuje se da kada se gol-razlika poveća za 1, prihod će se linearno i u prosjeku povećati za 0,41 milijuna funti što znači da svaki postignuti gol vrijedi u prosjeku 410 000 funti dok također svaki primljeni zgoditak vrijedi u prosjeku isto toliko, ali naravno u deficitarnom smislu. Vrijedi naglasiti da ovi iznosi vrijede samo za sezonu 2020/21 pošto Nogometni savezi zbog godišnje inflacije svake godine povećavaju izdatke za klubove pa je već iduće sezone 2021/22 prvak dobio 44 milijuna funti prihoda što je za 6 milijuna više nego promatrane 2020/21 sezone, a što su veći prihodi biti će i veća prosječna vrijednost za postignuti odnosno primljeni zgoditak.

To znači da ako je gol-razlika nekoga kluba npr. pet po procijenjenom regresijskom modelu očekuje se zarada od 22 milijuna funti, što je veoma reprezentativno uzevši u obzir da je klub sa gol-razlikom 8 zaradio 23 milijuna funti u promatranoj godini. Bio je to Leeds United.

$$\hat{y}_i = 19,95 + 0,41 \cdot 5 = 22$$

Ako neki klub sezonu završi sa gol-razlikom -22 očekivani prihod od saveza se po procijenjenom regresijskom modelu iznosi 10,93 milijuna funti što je veoma slično ostvarenom prihodu petnaestoplasiranog Southamptonu sa gol-razlikom -21 jer je taj klub ostvario prihod od 11 milijuna funti.

$$\hat{y}_l = 19,95 + 0,41 \cdot (-22) = 10,93$$

Prognoziranje vrijednosti prihoda, kako je navedeno, opravdano je jer je procijenjenim regresijskim modelom protumačeno 91,68% svih odstupanja (kako pokazuje koeficijent determinacije) što model čini vrlo reprezentativnim modelom.

Numeričkom analizom također je utvrđeno da je koeficijent linearne korelacije  $r = 0,9575$  što ukazuje na pozitivnu i jaku povezanost promatranih varijabli. Isti zaključak donesen je već i temeljem dijagrama rasipanja.

Standardna devijacija regresije izračunana po izrazu (26) iznosi 3,10 milijardi funti.

$$\sigma_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} = \sqrt{\frac{324,795}{20}} = 3,10$$

Koeficijent varijacije regresije izračunan po izrazu (27) je 15,56%

$$V_{\hat{y}} = \frac{\sigma_{\hat{y}}}{\bar{y}} \cdot 100 = \frac{3,10}{19,95} \cdot 100 = 15,56\%$$

Zaključuje se da je prosječno odstupanje stvarnih od regresijski (po modelu procijenjenih vrijednosti) prihoda 3,10 milijardi funti ili 15,56%. Zaključuje se da je i ovaj model zadovoljavajuće reprezentativan.

## 9. ZAKLJUČAK

Glavni zaključak rada jest da postoji jasna i jaka linearna povezanost između gol-razlike i broja bodova te između gol-razlike i prihoda na kraju sezone. Glavni razlog tomu jest činjenica da će npr. klub s pozitivnom gol-razlikom statistički imati više pobjeda nego klub s negativnom gol-razlikom, a što klub ima više pobjeda ima više bodova na tablici. Regresijskom analizom uočene su pravilnosti u odnosu navedenih varijabli te je kao takva iznimno korisna za nogometnu industriju jer otkriva vrlo precizno koliko postignuti pogodak u prosjeku znači prihoda od

nacionanog saveza za klub, odnosno koliko primljeni zgoditak osiromašuje klubsku blagajnu kao i koliko postignuti odnosno primljeni pogodak znači ostvarenih ili izgubljenih bodova. Bitno je istaknuti da bi se ovakva analiza trebala raditi prije početka svake sezone pošto svake sezone Nogometni savez povećava prihode za postignuti rezultat pa isto tako i postignuti i primljeni pogotci više odnosno manje vrijede.

Temeljem dostupnih podataka za sezonu 2020/2021 procijenjeni regresijski modeli su se pokazali kao vrlo reprezentativni te su korišteni za predviđanje broja ostvarenih bodova i za predviđanje prihoda ovisno o gol-razlici. Predviđanje bodova uz pomoć gol-razlike može biti ključno kada neki klub „lovi“ određenu poziciju ili zaostatak za nekim drugim klubom gdje se jasno može odrediti koliko je dovoljno postići pogodaka ili ih ne primiti da se postigne zacrtani cilj.

Treba naglasiti kako navedeni prihodi nisu jedini prihodi koji klubovi ostvaruju, ali svejedno čine solidan udio od u prosjeku 10 % do 15 % ukupnih prihoda klubova. Za neke nepopularne i novoosnovane klubova iz manjih mjesta ovi podaci bi bili poprilično ohrabrujući pošto udio prihoda od saveza u ukupnim prihodima takvih klubova iznosi puno više od gore spomenutih 10% do 15%.

## 10. PRILOZI

Linearna regresija broj 1 – odnos gol-razlike i bodova koje su klubovi ukupno ostvarili na kraju sezone 2020/2021

### SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0,969566233
R Square	0,940058681
Adjusted R Square	0,936728608
Standard Error	4,247842817
Observations	20

### ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	5093,754965	5093,755	282,293691	1,90484E-12
Residual	18	324,7950347	18,04417	6	
Total	19	5418,55			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	52,85	0,94984653	55,64057	1,34015E-21	50,85444649	54,8455535
X Variable 1	0,641288552	0,038168307	16,8016	1,90484E-12	0,561099914	0,72147719

### RESIDUAL OUTPUT

<i>Observation</i>	<i>Predicted Y</i>	<i>Residuals</i>	<i>Standard Residuals</i>
1	85,55571613	0,444283869	0,107456
2	71,447368	2,552632004	0,617391
3	69,52350234	-0,523502341	-0,12662
4	66,95834813	0,041651865	0,010074
5	64,39319393	1,606806071	0,388629
6	62,46932827	2,530671726	0,61208

7	67,59963669	-5,599636687	-1,35435
8	63,11061683	-2,110616825	-0,51048
9	57,98030841	1,019691587	0,246627
10	52,20871145	6,791288552	1,642572
11	58,62159696	-3,621596964	-0,87594
12	42,58938317	2,410616825	0,583043
13	42,58938317	2,410616825	0,583043
14	36,81778621	7,18221379	1,737122
15	39,38294042	3,617059583	0,874838
16	49,00226869	-8,00226869	-1,93546
17	38,74165187	0,258348135	0,062485
18	36,17649766	-8,176497659	-1,9776
19	26,55716938	-0,557169385	-0,13476
20	25,27459228	-2,274592282	-0,55014

---

Linearna regresija broj 2 – odnos gol-razlike i prihoda koje su klubovi zaradili od Engleskog nogometnog saveza na kraju sezone 2020/2021

SUMMARY  
OUTPUT

<i>Regression Statistics</i>	
Multiple R	0,957483993
R Square	0,916775597
Adjusted R Square	0,912152019
Standard Error	3,271244401
Observations	20

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	2121,831281	2121,831	198,2827	3,69358E-11
Residual	18	192,6187187	10,70104		
Total	19	2314,45			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	19,95	0,731472485	27,27375	4,3E-16	18,413233	21,486767
X Variable 1	0,41389472	0,02939324	14,08129	3,69E-11	0,3521418	0,4756476

RESIDUAL  
OUTPUT

<i>Observation</i>	<i>Predicted Y</i>	<i>Residuals</i>	<i>Standard Residuals</i>
1	41,05863071	-3,058630712	-0,96063
2	31,95294688	4,047053124	1,271061
3	30,71126272	3,288737284	1,032896
4	29,05568384	2,944316163	0,924724
5	27,40010496	2,599895043	0,816551
6	26,1584208	1,841579202	0,578386
7	29,46957856	-2,469578556	-0,77562
8	26,57231552	-2,572315518	-0,80789
9	23,26115776	-0,261157759	-0,08202
10	19,53610528	1,46389472	0,459767
11	23,67505248	-4,675052479	-1,4683

12	13,32768448	3,672315518	1,153367
13	13,32768448	1,672315518	0,525225
14	9,602632004	3,397367996	1,067014
15	11,25821088	-0,258210883	-0,0811
16	17,46663168	-7,466631681	-2,34505
17	10,84431616	-2,344316163	-0,73628
18	9,188737284	-3,188737284	-1,00149
19	2,980316486	1,019683514	0,320253
20	2,152527047	0,347472953	0,109131

---



## 11. LITERATURA

1. BeSoccer (2021.) What are the richest football clubs in the world (online). Dostupno na: [besoccer.com/new/what-are-the-richest-football-leagues-in-the-world-816834](https://besoccer.com/new/what-are-the-richest-football-leagues-in-the-world-816834) (14.6.2022.)
2. Čižmešija, M. i Kurnoga-Živadinović, N. (2012). Uvod u gospodarsku statistiku, Element. Zagreb.
3. Darren, W. (2021.) How much prize money every Premier League club made on their final position : Premier League prize money 2020/21 (online). Dostupno na [mirror.co.uk/sport/football/news/premier-league-club-prize-money-24168810](https://mirror.co.uk/sport/football/news/premier-league-club-prize-money-24168810) ( 7.5.2022.)
4. Flashcore (2021.) Football : Premier liga (online). Dostupno na: [flashscore.com/football/england/premier-league-2020-2021/](https://flashscore.com/football/england/premier-league-2020-2021/) (16.5.2022.)
5. FourFourTwo(2020.) Premier League 2020/21 preview: every club assessed ahead of the new season (online) Dostupno na: [fourfourtwo.com/features/premier-league-preview-2020-2021-season-every-club](https://fourfourtwo.com/features/premier-league-preview-2020-2021-season-every-club) (11.5.2022.)ž Rhett, L. (2021.) English Premier League History: The Origins And Formation (online). Dostupno na: [historyofsoccer.info/premier-league-founded](https://historyofsoccer.info/premier-league-founded) (19.5.2022.)
6. Ranđelović, D. (2021.) Athletic Panda : Most profitable sports leagues (online). Dostupno na : [apsportseditors.org/others/most-profitable-sports-leagues/](https://apsportseditors.org/others/most-profitable-sports-leagues/)
7. Stedy (2021.) Statističko zaključivanje : jednostavna linearna regresija (online). Dostupno na [stedy.hr/statisticko-zakljucivanje/jednostavna-linearna-regresija](https://stedy.hr/statisticko-zakljucivanje/jednostavna-linearna-regresija) (4.5.2022.)
8. Sofascore (2021). Premier League live score table (online). Dostupno na [sofascore.com/tournament/football/england/premier-league/17](https://sofascore.com/tournament/football/england/premier-league/17) (8.5.2022.)
9. Šošić I. (2006) Primijenjena statistika. Školska knjiga, Zagreb (odabrana poglavlja)
10. Šošić I., Serdar V. (2002). Uvod u statistiku, Školska knjiga, Zagreb.
11. Transfermarkt (2021) Squad of Manchester City : Overview (online). Dostupno na : [transfermarkt.com/manchester-city/startseite/verein/281](https://transfermarkt.com/manchester-city/startseite/verein/281) (13.5.2022.)
12. Transfermarkt (2021.) Kevin de Bruyne : Profile (online). Dostupno na: [transfermarkt.com/kevin-de-bruyne/profil/spieler/88755](https://transfermarkt.com/kevin-de-bruyne/profil/spieler/88755) (16.5.2022.)



## POPIS TABLICA :

Tablica 1. Odnos koeficijenta determinacije i korelacije

Tablica 2. Ime kluba, pozicija na tablici, gol-razlika, prihod te ostvareni bodova klubova

## POPIS GRAFIKONA:

Grafikon 1. Dijagram rasipanja za gol-razliku i bodove koje su klubovi ukupno ostvarili na kraju sezone 2020/2021

Grafikon 2. Dijagram rasipanja za gol-razliku i prihoda koje su klubovi ostvarili od Engleskog nogometnog saveza na kraju sezone 2020/2021