

SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
ODJEL ZA MATEMATIKU

Sveučilišni preddiplomski studij Matematika i računarstvo

**ANALIZA POSJEĆENOSTI UTAKMICA PRVE
HRVATSKE NOGOMETNE LIGE**

Seminarski rad iz kolegija: Statistički praktikum

Matej Kolak

Osijek, 2022.

Sadržaj

1. Uvod	1
2. Opis baze podataka	2
3. Analiza	4
3.1 Posjećenost utakmica s obzirom na sudionike natjecanja	4
3.1.1 Broj gledatelja s obzirom na domaću momčad	4
3.1.2. Broj gledatelja s obzirom na gostujuću momčad	6
3.1.3. Broj gledatelja s obzirom na međusobne utakmice dvaju momčadi	8
3.2. Posjećenost utakmica s obzirom na kompetitivne čimbenike	10
3.2.1. Rast broja gledatelja s obzirom na pobjede	10
3.2.2. Broj gledatelja s obzirom na elo rejting kluba	11
3.3. Posjećenost utakmica s obzirom na vremenske čimbenike	13
3.3.1. Ovisnost broja gledatelja o godišnjem dobu	13
3.3.2. Broj gledatelja s obzirom na dan u tjednu	14
4. Zaključak	17
5. Izvori	18

1. Uvod

Prva Hrvatska nogometna liga, poznatija po kratici HNL, najviša je razina natjecanja u hrvatskom nogometu. Svake godine pobjednik natjecanja odnosi titulu državnog prvaka u nogometu i kao takav predstavlja Hrvatsku u europskim natjecanjima.

Prva sezona natjecanja odigrala se 1992. godine, a sve do danas odigrana je ukupno 31 sezona, zaključno sa netom održanom sezonom 2021/22. Natjecanje se kroz povijest održavalo u različitim formatima, prema kojima je u ligi sudjelovalo od 10, 12, 16 pa sve do najviše 22 različita kluba.

Današnji format natjecanja ustrojen je prije 9 godine i sastoji se od 10 klubova koji međusobno igraju po 4 utakmice, 2 na domaćem i 2 na gostujućem terenu. To osigurava ukupno 180 odigranih utakmica po sezoni. U ovom formatu posljednjeplasirani klub ispada iz natjecanja, a njegovo mjesto u idućoj sezoni preuzima najbolje plasirani klub druge razine natjecanja.

Tema ovog seminara je istraživanje posjećenosti utakmica ovog natjecanja s ciljem utvrđivanja čimbenika koji utječu na istu.

2. Opis baze podataka

Seminarski rad baziran je na podacima prikupljenim sa službene web stranice natjecanja. Prikupljeni su podaci o 1260 utakmica s 15 obilježja, počevši sa sezonom 2015/16 i zaključno sa sezonom 2021/22. Službeni podaci o broju gledatelja na utakmicama prije sezone 2015/16 nisu dostupni i stoga baza ne sadrži sve utakmice od samog početka natjecanja. Važno je napomenuti da je u sezoni 2020/21 oko 90% utakmica odigrano bez gledatelja zbog COVID restrikcija. U ostalim sezonama postoji još nekoliko utakmica koje su se zbog disciplinskih mjera nogometnog saveza odigrale bez gledatelja. Posjećenost tih utakmica neće biti korištena u istraživanju s obzirom da razlog nedolaska gledatelja na utakmicu nije uzrokovan manjkom interesa. S druge strane, rezultati tih utakmica koristiti će se za izračun kompetitivnih faktora.

Elo sustav

Osim službenih podataka, svakoj utakmici u bazi dodana su i obilježja vezana uz elo sustav. Elo sustav koristi se relativnu procjenu trenutne kompetitivne “snage” nekog kluba. Svakom klubu pridružuje se rejting koji predstavlja njegovu “snagu”. U slučaju pobjede, taj rejting raste, dok u slučaju poraza pada. Međutim, promjena rejtinga ne ovisi isključivo o rezultatu već i o snazi protivnika pa se tako pobjede protiv jačih protivnika vrednuju značajno više nego pobjede protiv slabijih protivnika. Elo sustav koristi sustav matematičkih jednadžbi za izračune rasta, odnosno pada rejtinga i te jednadžbe ovise isključivo o razlici između rejtinga timova, broju postignutih golova te statusu domaćina.

Opis varijabli

1. **Season** - kvalitativna nominalna varijabla koja označava sezonu odigravanja utakmice
2. **Weekday** - kvalitativna ordinalna varijabla, označava dan u tjednu odigravanja utakmice
3. **Day** - numerička diskretna varijabla - označava dan u mjesecu. Zajedno s **Month** i **Year** tvori datum odigravanja utakmice. Datum je rastavljen na ove tri varijable iz tehničkih razloga.
4. **Month** - numerička diskretna varijabla - označava mjesec u godini.
5. **Year** - numerička diskretna varijabla - označava godinu odigravanja.

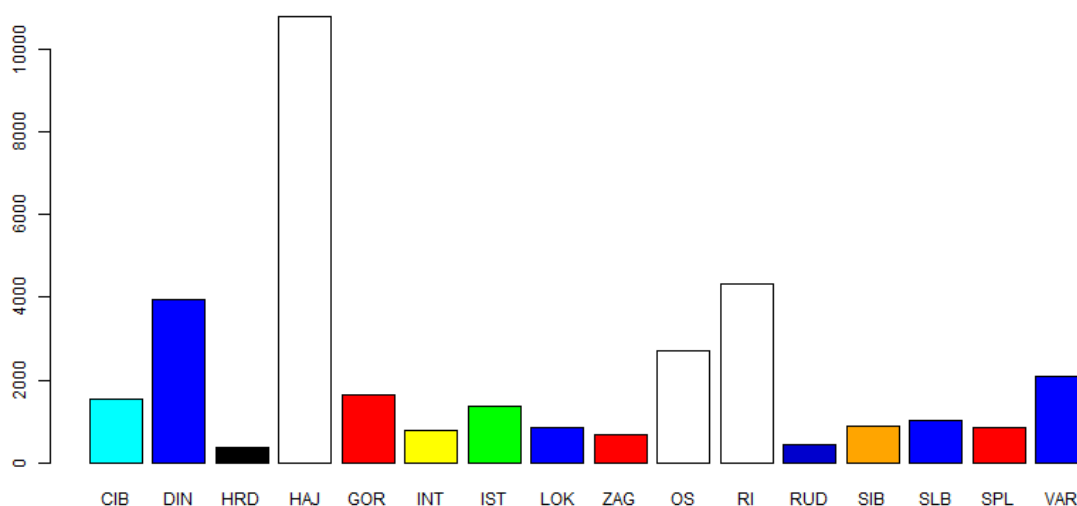
6. **Home team** - kvalitativna nominalna varijabla - označava prvu momčad koja sudjeluje u utakmici. To je ujedno i momčad na čijem se terenu odigrava utakmica.
7. **Away team** - kvalitativna nominalna varijabla - označava drugu momčad koja sudjeluje u utakmici.
8. **Home elo** - numerička diskretna varijabla - označava elo rating domaće momčadi prije odigravanja utakmice.
9. **Away elo** - numerička diskretna varijabla - označava elo rating gostujuće momčadi prije odigravanja utakmice.
10. **Home goals** - numerička diskretna varijabla - označava broj golova koje je postigla domaća momčad na kraju utakmice. Zajedno s **Away goals** služi za utvrđivanje konačnog rezultata utakmice.
11. **Away goals** - numerička diskretna varijabla - označava broj golova koje je postigla gostujuća momčad na kraju utakmice.
12. **Home odds** - numerička neprekidna varijabla - označava vjerojatnost pobjede domaće momčadi. **Home odds**, **Draw odds** i **Away odds** u sumi moraju iznositi točno 1. Vjerojatnost pobjede izračunata je pomoću matematičkih modela elo sustava koji u obzir uzima razliku u elo rejtinzima prije utakmice i prednost domaćeg terena.
13. **Draw odds** - numerička neprekidna varijabla - označava vjerojatnost neriješenog rezultata utakmice.
14. **Away odds** - numerička neprekidna varijabla - označava vjerojatnost pobjede gostujuće momčadi.
15. **Attendance** - numerička diskretna varijabla - označava broj gledatelja prisutnih na utakmici.

3. Analiza

3.1 Posjećenost utakmica s obzirom na sudionike natjecanja

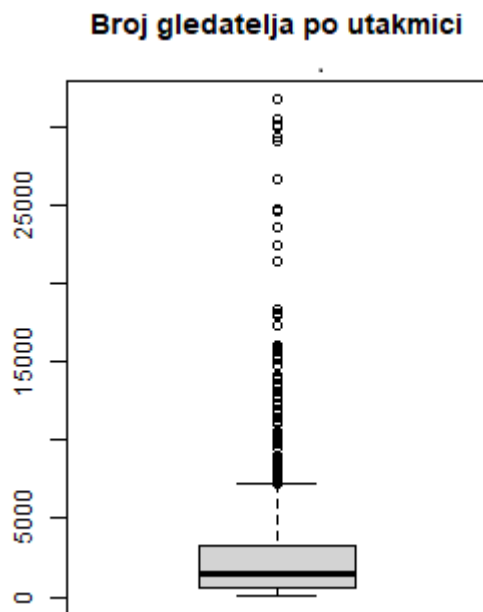
3.1.1 Broj gledatelja s obzirom na domaću momčad

Posljednjih 7 godina u Prvoj Hrvatskoj nogometnoj ligi nastupilo je 16 različitih klubova. Svaki od tih klubova dao je određeni doprinos ukupnoj posjećenosti utakmica.



Slika 1: Prosječni broj gledatelja domaćih utakmica svakog kluba

Ono što se prvo može primjetiti jest da po broju gledatelja domaćih utakmica uvjerljivo prednjači Hajduk sa prosječno 10800.69 gledatelja po utakmici. Slijede ga Rijeka sa 4342.39, Dinamo sa 3951.94 i Osijek sa 2717.64 gledatelja po utakmici. S druge strane, najmanju gledanost imaju zagrebački klubovi. Domaće utakmice Hrvatskog Dragovoljca prosječno prati 358.11 gledatelja, Rudeša 453.42, Zagreba 687.11, a Lokomotive 865.33 gledatelja po utakmici. Uz zagrebačke klubove, u ovu skupinu spada i Inter Zaprešić sa prosječno 773.56 gledatelja po utakmicu. Nesrazmjer podataka dolazi do izražaja kada usporedimo medijan i aritmetičku sredinu broja gledatelja po utakmici. Aritmetička sredina uzorka svih utakmica iznosi 2776.1 gledatelj po utakmici, dok je medijan uzorka 1456 gledatelja po utakmici.



Kutijasti dijagram broja gledatelja po utakmici sugerira da je maksimum uzorka 7181, a sve vrijednosti iznad smatraju se stršećim. Kada izračunamo da je 0.4-kvantil broja gledatelja Hajdukovih domaćih utakmica 7790, vidimo da u čitavom uzorku čak 60% Hajdukovih domaćih utakmica, njih čak 61, zapravo spada u stršeće vrijednosti. Kako bi ukazali na nesrazmjer između posjećenosti utakmica s obzirom na klub domaćin, provodimo sljedeća 2 testa.

Slika 2: Kutijasti dijagram broja gledatelja

Dinamo i Hajduk povijesno su 2 najuspješnija i najpopularnija kluba u Hrvatskoj. Smatra se da gledatelji koji posjećuju domaće utakmice ta dva kluba čine više od 50% ukupnog broja gledatelja u čitavoj populaciji.

Testiramo sljedeće hipoteze:

H_0 : Proporcija ukupnog broja gledatelja domaćih utakmica Dinama i Hajduka u čitavoj populaciji jednaka je 0.5

H_1 : Proporcija ukupnog broja gledatelja domaćih utakmica Dinama i Hajduka u čitavoj populaciji veća je od 0.5

Provodimo binomni test i njime dobivamo $p\text{-vrijednost} = 2.2e-16 < 0.05$ pa na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo alternativnu hipotezu H_1 . Procjena proporcije je 0.5134964. Ovim testom pokazali smo da gledatelji domaćih utakmica dvaju najpopularnijih klubova čine više od 50% publike na stadionima, dok preostalih 14 klubova privlači ostatak publike.

Kada Hajduku i Dinamu pridružimo sljedeća 2 najpopularnija kluba, Osijek i Rijeku, dolazimo do tvrdnje da gledatelji koji posjećuju domaće utakmice ta četiri kluba zajedno čine više od 75% ukupnog broja gledatelja u čitavoj populaciji.

Testiramo sljedeće hipoteze:

H_0 : Proporcija ukupnog broja gledatelja domaćih utakmica Dinama, Hajduka, Rijeke i Osijeka u čitavoj populaciji jednaka je 0.75

H_1 : Proporcija ukupnog broja gledatelja domaćih utakmica Dinama, Hajduka, Rijeke i Osijeka u čitavoj populaciji veća je od 0.75

Provodimo binomni test i njime dobivamo $p\text{-vrijednost} = 2.2e-16 < 0.05$ pa na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo alternativnu hipotezu H_1 . Procjena proporcije je 0.7689936. Ovim testom pokazali smo da su gledatelji domaćih utakmica četiriju najpopularnijih hrvatskih klubova, koji čine jednu četvrtinu klubova u populaciji, zaslužni za više od tri četvrtine ukupne posjećenosti utakmica Prve Hrvatske nogometne lige.

3.1.2. Broj gledatelja s obzirom na gostujuću momčad

Često posjećenost utakmica ne ovisi isključivo o domaćem klubu, već i o ekipi koja dolazi u goste. Čimbenici koji čine najveću razliku u ovoj kategoriji trebali bi ponovo biti Dinamo i Hajduk. Navijačke skupine Dinama i Hajduka najbrojnije su u Hrvatskoj. Oba kluba također su popularna diljem države i imaju velik broj navijača izvan gradova iz kojih potječu. Provesti ćemo dva testa koji će ispitati kako ta dva kluba utječu na očekivani broj gledatelja u gostujućim utakmicama.

Testiramo sljedeće hipoteze:

H_0 : Očekivani broj gledatelja na utakmicama u kojima je gostujuća momčad Dinamo jednak je očekivanom broju gledatelja na utakmicama u kojima ne nastupa Dinamo.

H_1 : Očekivani broj gledatelja na utakmicama u kojima je gostujuća momčad Dinamo veći je od očekivanog broja gledatelja na utakmicama u kojima ne nastupa Dinamo.

Provodimo Welchov t-test i njime dobivamo $p\text{-vrijednost} = 5.525e-05 < 0.05$ pa na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo alternativnu hipotezu H_1 . Procjena očekivanja broja gledatelja na utakmicama u kojima gostuje momčad Dinama je 5191.019, dok je očekivani broj gledatelja na ostalim utakmicama 2326.087. Isti test provodimo i za gostujuće utakmice Hajduka.

Testiramo sljedeće hipoteze:

H_0 : Očekivani broj gledatelja na utakmicama u kojima je gostujuća momčad Hajduk jednak je očekivanom broju gledatelja na utakmicama u kojima ne nastupa Hajduk.

H_1 : Očekivani broj gledatelja na utakmicama u kojima je gostujuća momčad Hajduk veći je od očekivanog broja gledatelja na utakmicama u kojima ne nastupa Hajduk.

Provodimo Welchov t-test i njime dobivamo $p\text{-vrijednost} = 9.008e-11 < 0.05$ pa na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo alternativnu hipotezu H_1 . Procjena očekivanja broj gledatelja na utakmicama u kojima gostuje momčad Hajduka je 4360.991, dok je očekivani broj gledatelja na ostalim utakmicama 1635.121. Posljednja dva testa pokazala su da gostovanja Dinama i Hajduka utječu na posjećenost utakmica.

Osim Hajduka i Dinama, ranije smo uvidjeli da Rijeka i Osijek također imaju iznadprosječan broj gledatelja na svojim domaćim utakmicama. Gradovi Osijek i Rijeka imaju manji broj stanovnika od Splita i Zagreba te njihovi istoimeni klubovi nemaju veliki broj navijača izvan svojih gradova. Sljedeći testovi ispituju utječu li gostovanja Osijeka i Rijeke na očekivani broj gledatelja. Testiramo sljedeće hipoteze:

H_0 : Očekivani broj gledatelja na utakmicama u kojima je gostujuća momčad Osijek jednak je očekivanom broju gledatelja na utakmicama u kojima ne nastupa Osijek.

H_1 : Očekivani broj gledatelja na utakmicama u kojima je gostujuća momčad Osijek veći je od očekivanog broja gledatelja na utakmicama u kojima ne nastupa Osijek.

Provodimo Welchov t-test i njime dobivamo $p\text{-vrijednost} = 0.5938 > 0.05$ pa na razini značajnosti 0.05 ne odbacujemo H_0 . Nemamo dovoljno dokaza da gostujuće utakmice Osijeka imaju veću posjećenost od utakmica u kojima ne nastupa Osijek. Procjena očekivanog broja gledatelja na gostujućim utakmicama Osijeka je 2695.852, a za ostale utakmice 2793.544. Idući test provodimo s obzirom na gostujuće utakmice Rijeke.

Testiramo sljedeće hipoteze:

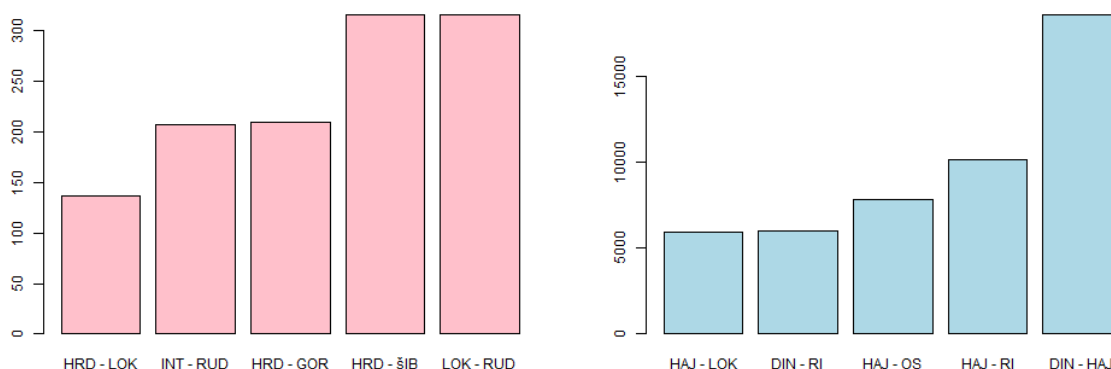
H_0 : Očekivani broj gledatelja na utakmicama u kojima je gostujuća momčad Rijeka jednak je očekivanom broju gledatelja na utakmicama u kojima ne nastupa Rijeka.

H_1 : Očekivani broj gledatelja na utakmicama u kojima je gostujuća momčad Rijeka veći je od očekivanog broja gledatelja na utakmicama u kojima ne nastupa Rijeka.

Provodimo Welchov t-test i njime dobivamo $p\text{-vrijednost} = 0.0347 < 0.05$ pa na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo alternativnu hipotezu H_1 . Procjena očekivanja broj gledatelja na utakmicama u kojima gostuje momčad Rijeke je 3351.552, dok je očekivani broj gledatelja na ostalim utakmicama 2509.768.

3.1.3. Broj gledatelja s obzirom na međusobne utakmice dvaju momčadi

Od sezone 2015/2016, ukupno 98 različitih parova klubova je odigralo međusobne utakmice. Pritom je svaki par, s obzirom na format natjecanja, odigrao minimalno 4 međusobne utakmice.



Slika 3: Parovi klubova sa najmanjim(lijevo) i najvećim(desno) prosječnim brojem gledatelja u međusobnim susretima.

Prema lijevoj grafici(Slika 3), vidimo da su čak 3 para u kojima sudjeluje Hrvatski Dragovoljac su među 5 prosječno najslabije posjećenih utakmica u uzorku. Ističe se gradski dvoboj Hrvatskog Dragovoljca i Lokomotive sa prosječnih 136.5 gledatelja po utakmici. S druge strane, u 4 od 5 parova s prosječno najvećim brojem gledatelja sudjeluje Hajduk. Kao uvjerljivo

najposjećeniji par u čitavom uzorku ističe se dvoboj Hajduka i Dinama sa prosječno 18631.875 gledatelja po utakmici. Najmanji broj gledatelja u uzorku utakmica Hajduka i Dinama je 1967 gledatelja na utakmici odigranoj 12.7.2020. Želimo vidjeti koliko je utakmica u populaciji svih utakmica slabije posjećeno od najslabije posjećene utakmice Hajduka i Dinama. Procjena 0.6-kvantila populacije je 1648.2, a procjena 0.65-kvantila 2012. Testiramo sljedeće hipoteze:

$$\begin{aligned} H_0: & 0.6\text{-kvantil broja gledatelja na utakmicama jednak je 1967.} \\ H_1: & 0.6\text{-kvantil broja gledatelja na utakmicama manji je od 1967.} \end{aligned}$$

Provođenjem testa o kvantilima dobivamo $p\text{-vrijednost} = 0.0009846 < 0.05$ pa na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo alternativnu hipotezu H_1 . Time smo zaključili da je i najslabije posjećena utakmica Hajduka i Dinama posjećenija od 60% svih odigranih utakmica. Važna stavka kod te utakmice jest da je tako niska posjećenost uzrokovana ograničenjima broja navijača na stadionu u razdoblju COVID-a. Stoga kada bismo tražili najslabije posjećenu utakmicu u normalnim uvjetima bez ikakvih ograničenja, došli bismo do utakmice odigrane 5.12.2021. na koju je došlo 9974 gledatelja. Procjena 0.94-kvantila je 7900.18, a 0.97-kvantila 11887.49. Testiramo sljedeće hipoteze:

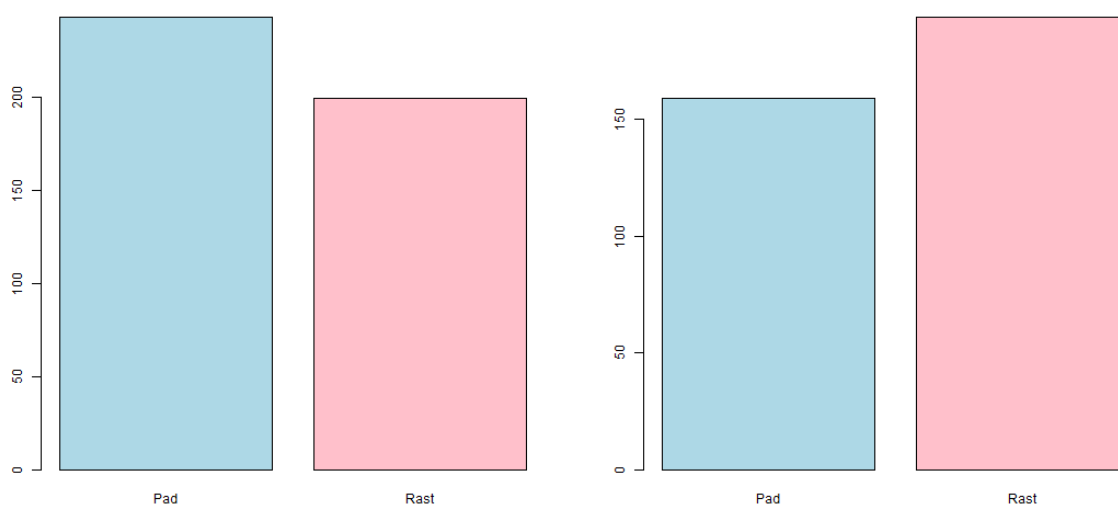
$$\begin{aligned} H_0: & 0.94\text{-kvantil broja gledatelja na utakmicama jednak je 9974.} \\ H_1: & 0.94\text{-kvantil broja gledatelja na utakmicama manji je od 9974.} \end{aligned}$$

Provodimo test o kvantilima i dobivamo $p\text{-vrijednost} = 0.004543 < 0.05$. Na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo H_1 . Ovim testom dokazali smo da je najslabije posjećena utakmica Hajduka i Dinama u normalnim okolnostima posjećenija od 94% svih odigranih utakmica.

3.2. Posjećenost utakmica s obzirom na kompetitivne čimbenike

3.2.1. Rast broja gledatelja s obzirom na pobjede

U ovom dijelu seminara se pitamo utječu li pobjede kluba na porast broja gledatelja u domaćim utakmicama kluba. Da bismo ispitali ovu tezu, za svaki klub stvoren je novi, zaseban uzorak koji je stvoren na sljedeći način. Promatra se uzorak domaćih i gostujućih utakmica jednog kluba, sortiranih po datumima. Za svaku utakmicu koju je klub pobijedio, neovisno o statusu domaćina, bilježi se posljednja domaća utakmica koju je klub igrao prije pobjede (koja ujedno može biti i utakmica kojoj je klub pobijedio) i prva sljedeća domaća utakmica. Proces se vrši na taj način jer klubovi rijetko igraju dvije uzastopne domaće utakmice. U uzorak zapisujemo 1 ukoliko je broj gledatelja u utakmici poslije pobjede veći od broja gledatelja u utakmici prije pobjede, a 0 u suprotnom.



Slika 4: Frekvencija trenda pada i rasta nakon poraza(lijevo) i nakon pobjede(desno)

Želimo testirati proporciju jedinica u gore opisanom uzorku.

Testiramo sljedeće hipoteze:

H_0 : Proporcija trenda rasta broja gledatelja nakon pobjede jednaka je 0.5

H_1 : Proporcija trenda rasta broja gledatelja nakon pobjede veća je od 0.5

Provodimo binomni test i dobivamo $p\text{-vrijednost} = 0.0351 < 0.05$. Na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo H_1 . Procjena proporcije je

0.5495751. Ovim testom pokazali smo da pobjede imaju utjecaj na porast broja gledatelja.

Sljedeći uzorak generiramo na analogan način. Umjesto pobjeda, sada promatramo utakmice u kojima je klub doživio poraz. U novonastali uzorak bilježimo 1 ukoliko je broj gledatelja na utakmici poslije poraza veći od broja gledatelja na utakmici prije poraza, a 0 u suprotnom.

Testiramo sljedeće hipoteze:

H_0 : Proporcija trenda pada broja gledatelja nakon poraza jednaka je 0.5

H_1 : Proporcija trenda pada broja gledatelja nakon poraza veća je od 0.5

Provodimo binomni test i dobivamo $p\text{-vrijednost} = 0.02035 < 0.05$. Na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo H_1 . Procjena proporcije je 0.5497738. Testom smo dokazali da poraz kluba utječe na smanjenje broja gledatelja.

3.2.2. Broj gledatelja s obzirom na elo rejting kluba

Elo rejting služi za ocjenu kompetitivne snage nekog kluba u određenom trenutku. Ono što se može naslutiti jest da bi kompetitivnije i uspješnije momčadi trebale privući više publike.

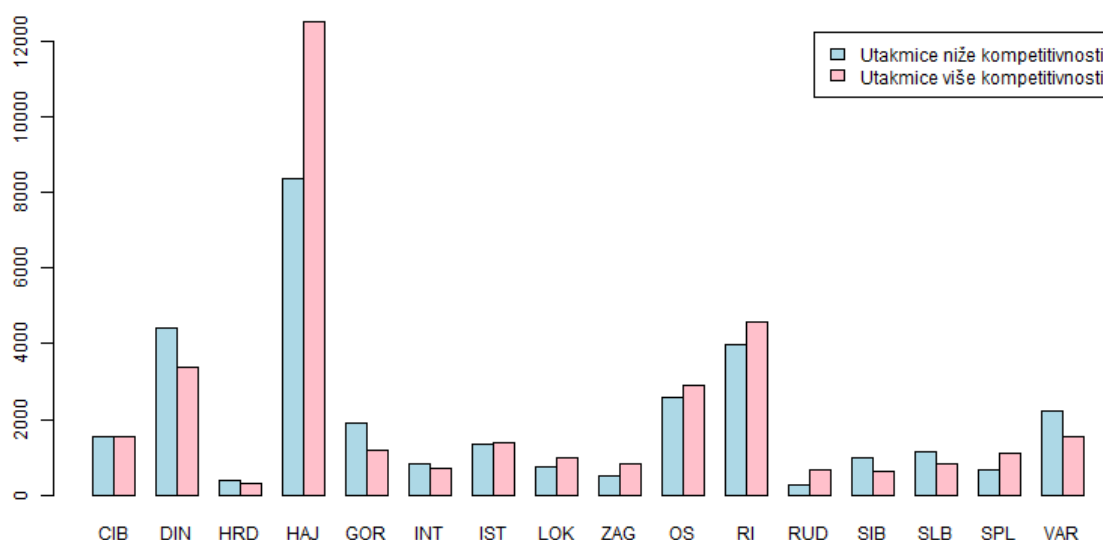
Za sljedeći test generiramo dva uzorka. Za svaki klub provodimo postupak u kojem njegove domaće utakmice dijelimo na uzorak domaćih utakmica u kojima je klub bio kompetitivniji i uzorak domaćih utakmica u kojima je klub bio manje kompetitivan. Prvi uzorak koji ćemo koristiti u testu biti će unija uzoraka utakmica više kompetitivnosti svih klubova, a drugi unija uzoraka utakmica niže kompetitivnosti svih klubova. Postupak kojim domaće utakmice nekog kluba dijelimo s obzirom na kompetitivnost je sljedeći. Iz uzorka svih utakmica kluba računamo njegov medijan elo rejtinga. Za svaku domaću utakmicu kluba uzimamo elo rejting s kojim je klub ušao u utakmicu i uspoređujemo ga s ranije izračunatim medijanom elo rejtinga. Utakmice u koje klub ulazi s većim elo rejtingom od medijana svrstavamo u utakmice više kompetitivnosti, a ostale utakmice svrstavamo u uzorak niže kompetitivnosti.

Testiramo sljedeće hipoteze:

H_0 : Očekivani broj gledatelja na utakmicama više kompetitivnosti klubova domaćina jednak je očekivanom broju gledatelja na utakmicama niže kompetitivnosti klubova domaćina.

H_1 : Očekivani broj gledatelja na utakmicama više kompetitivnosti klubova domaćina veći je od očekivanog broja gledatelja na utakmicama niže kompetitivnosti klubova domaćina.

Pokrećemo Welchov t-test i dobivamo $p\text{-vrijednost} = 0.000642 < 0.05$ pa na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo H_1 . Ovime smo pokazali da kompetitivnost klubova utječe na broj gledatelja. Procjena očekivanog broja gledatelja utakmica više kompetitivnosti je 3206.292, a za utakmice niže kompetitivnosti procjena je 2409.706.

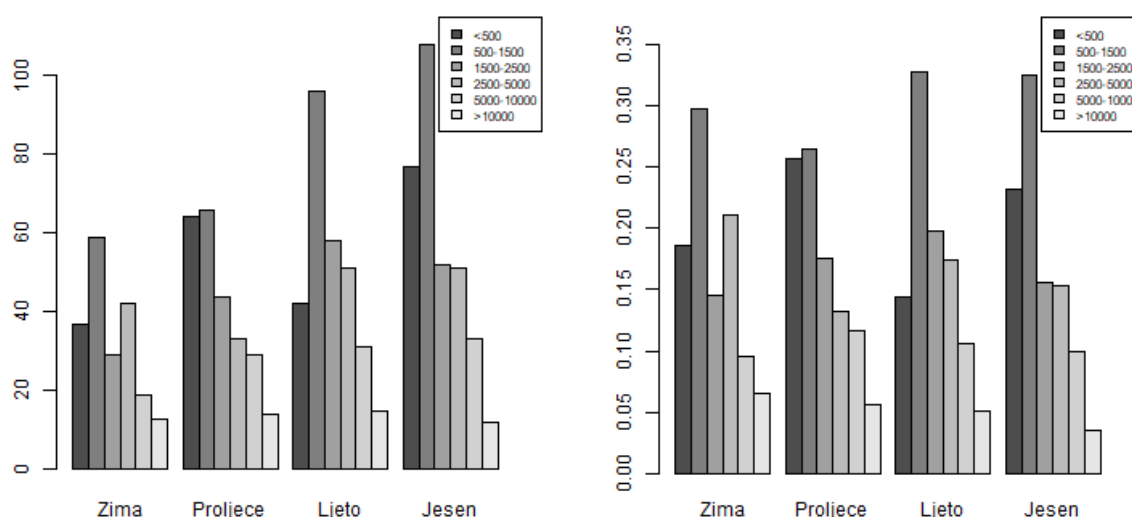


Slika 5: Prosječni broj gledatelja domaćih utakmica klubova s obzirom na razinu kompetitivnosti

3.3. Posjećenost utakmica s obzirom na vremenske čimbenike

3.3.1. Ovisnost broja gledatelja o godišnjem dobu

U javnosti se često kritizira stanje u kojem se nalaze hrvatski stadioni. Stadioni su često zastarjeli, neodržavani, nenatkriveni ili djelomično nenatkriveni. Stječe se dojam da zbog niske razine komfora, manjka zaštite u slučaju nepovoljnih vremenskih uvjeta i niskih temperatura gledatelji slabije posjećuju utakmice u razdoblju zime i kasne jeseni. Provodimo dva testa kako bismo testirali te slutnje. Za potrebe prvog testa, generiramo četiri uzorka u kojima utakmice dijelimo s obzirom na godišnje doba u kojem su odigrane. Dodatno, utakmice dijelimo na 6 kategorija s obzirom na broj gledatelja.



Slika 6: Tablica frekvencija(lijevo) i tablica relativnih frekvencija(desno) kategorija broja gledatelja na utakmici s obzirom na godišnje doba

Iz lijeve grafike možemo uočiti da se općenito najviše utakmica odigrava tijekom jeseni, a najmanje tijekom zime što je uzrokovano pauzom zimskom pauzom prvenstva koje uobičajeno traje između 30 i 45 dana. Kada promotrimo desnu grafiku, možemo primjetiti da se distribucije uzoraka ne razlikuju u prevelikoj mjeri. Želimo testirati postoji li razlika u distribucijama ova četiri uzorka.

Testiramo sljedeće hipoteze:

H_0 : Distribucije kategorija broja gledatelja jednake su po godišnjem dobu u kojem je utakmica odigrana

H_1 : Distribucije kategorija broja gledatelja nisu jednake po godišnjem dobu u kojem je utakmica odigrana

Provodimo χ^2 -test o homogenosti i dobivamo p -vrijednost = 0.09552 < 0.05 pa na razini značajnosti 0.05 ne odbacujemo H_0 . Nemamo dovoljno dokaza da su distribucije različite, stoga ne možemo tvrditi da postoji značajna razlika u broju gledatelja na utakmicama s obzirom na godišnje doba.

Ipak, želimo promatrati slutnju i iz drugog kuta. S obzirom da nisu dostupni službeni podaci o vremenskim uvjetima i temperaturi za svaku utakmicu, utakmice ćemo podijeliti s obzirom na datum odigravanja, odnosno da li je utakmica odigrana u razdoblju tipično nepovoljnih vremenskih uvjeta, tojest u razdoblju koje počinje 1.12., a završava 21.3. Promatramo očekivani broj gledatelja na utakmicama odigranim u tom razdoblju, naspram očekivanog broja gledatelja na svim ostalim utakmicama.

Testiramo sljedeće hipoteze:

H_0 : Očekivani broj gledatelja na utakmicama odigranim u razdoblju nepovoljnih vremenskih uvjeta jednak je očekivanom broju gledatelja u ostalim utakmicama.

H_1 : Očekivani broj gledatelja na utakmicama odigranim u razdoblju nepovoljnih vremenskih uvjeta manji je od očekivanog broju gledatelja u ostalim utakmicama..

Provodimo Welchov t-test i dobivamo p -vrijednost = 0.6937 < 0.05 pa na razini značajnosti 0.05 ne odbacujemo H_0 . Nemamo dovoljno dokaza da je broj gledatelja manji u razdoblju nepovoljnih uvjeta.

3.3.2. Broj gledatelja s obzirom na dan u tjednu

Nogometne utakmice najčešće se odigravaju vikendom i petkom. Ipak, veliki broj utakmica zaostalih utakmica svoj termin pronalaze među radnim danima. Zbog raznih životnih obaveza gledateljima je teže pronaći vrijeme za odlazak na nogometnu utakmicu preko tjedna. Iz istog razloga putovanja su nepraktična za

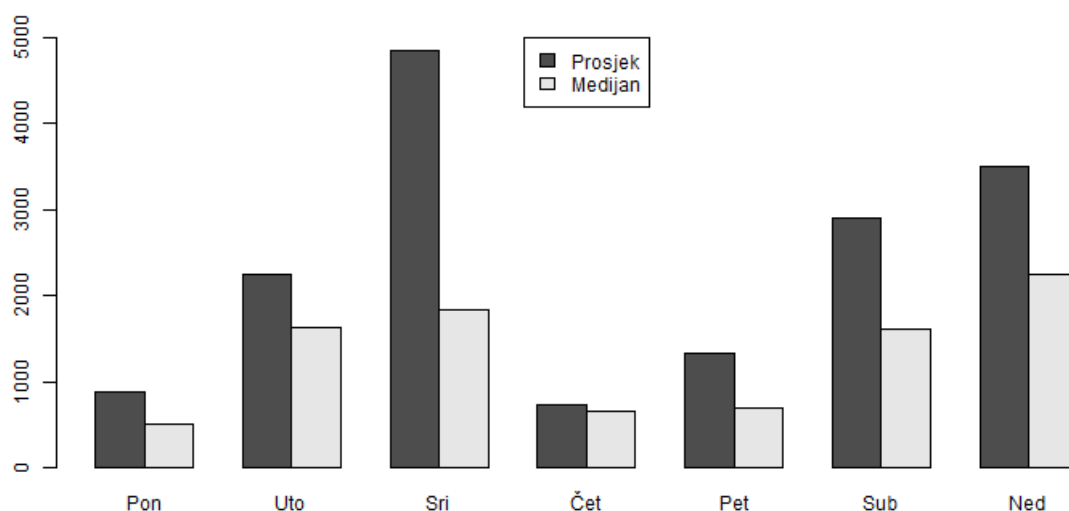
gostujuće navijače. Ta dva razloga trebala bi biti dovoljna za stvaranje razlike u broju gledatelja na utakmicama koje se igraju vikendom i onih koje se igraju radnim danom.

Testiramo sljedeće hipoteze:

H_0 : Očekivani broj gledatelja na utakmicama odigranim radnim danom jednak je očekivanom broju gledatelja na utakmicama odigranim vikendom

H_1 : Očekivani broj gledatelja na utakmicama odigranim radnim danom manji je od očekivanog broja gledatelja na utakmicama odigranim vikendom

Provodimo Welchov t-test i dobivamo $p\text{-vrijednost} = 5.886e-09 < 0.05$ pa na razini značajnosti 0.05 odbacujemo H_0 i prihvaćamo H_1 . Time smo dokazali da utakmice odigrane vikendom privlače više publike. Procjena očekivanja broja gledatelja za utakmice odigrane radnim danom je 1753.179, a za utakmice odigrane vikendom 3186.874.



Slika 7: Aritmetička sredina i medijan broja gledatelja s obzirom na dan odigravanja utakmice

Iz grafičkog prikaza (Slika 7) dalo bi se naslutiti da su utakmice odigrane srijedom popularnije od utakmica subotom i nedjeljom. Prosjek ipak ne otkriva stvarno stanje s obzirom da se srijedom odigralo ukupno 37 utakmica, a igrom slučaja čak 6 od tih 37 bilo je susret Dinama i Hajduka. Nedjeljom je odigrano daleko više, čak 415 utakmica i stoga je prosjek neminovno manji. Ipak, da bismo dokazali da ta razlika nije značajna provodimo test.

Testiramo sljedeće hipoteze:

H_0 : Očekivani broj gledatelja na utakmicama odigranim srijedom jednak je očekivanom broju gledatelja na utakmicama odigranim nedjeljom

H_1 : Očekivani broj gledatelja na utakmicama odigranim srijedom veći je od očekivanog broja gledatelja na utakmicama odigranim nedjeljom

Provodimo Welchov t-test i dobivamo $p\text{-vrijednost} = 0.08393 < 0.05$ pa na razini značajnosti 0.05 ne odbacujemo H_0 . Nemamo dovoljno dokaza da je očekivani broj gledatelja na nedjeljnim utakmicama manji od broja gledatelja utakmica srijedom.

4. Zaključak

Na osnovu analize provedene u ovom seminaru, došli smo do nekoliko zaključaka vezanih uz posjećenost utakmica Prve Hrvatske nogometne lige.

Već i prije same analize bili su poznati glavni akteri natjecanja, Dinamo, Hajduk, Rijeka i Osijek. Otprije je poznato da njihove utakmice privlače veći broj gledatelja, no analizom je dobiven uvid u to koliki je zapravo omjer publike naspram ostatka Hrvatske. Time je jasno da važnu ulogu igraju tradicija tih klubova i veličina gradova iz kojih potječu. Rijeka i Osijek potječu iz gradova slične veličine i kroz povijest su glasili kao treći i četvrti najveći klub u Hrvatskoj. Ipak, za razliku od Osijeka, Rijeka je u posljednjih 10 godina ostvarila značajne uspjehe, prvenstveno osvajanje naslova prvaka države 2017. godine. Time smo dobili da danas Rijeka ima drugu najveću posjećenost domaćih utakmica u Hrvatskoj, a istovremeno i privlači gledatelje na svojim gostovanjima, za razliku od Osijeka. Na primjeru Rijeke možemo reći da uspjesi i kompetitivnost kluba uistinu imaju utjecaj na publiku.

Kompetitivni čimbenici koje smo u daljnjem dijelu seminara obradili također su pokazali da publika u Hrvatskoj srazmjerno nagrađuje klubove svojim prisustvom ovisno o njihovoj uspješnosti.

Unatoč predrasudama javnosti, može se zaključiti da nepovoljni vremenski uvjeti još uvijek ne sprječavaju publiku od dolaska na stadione.

5. Izvori

- <https://hnl.hr/povijest/rezultati-i-poretdi/>
- https://en.wikipedia.org/wiki/Croatian_Football_League
- <http://clubelo.com/CRO/Games>