

Hokej v številkah in grafih

Kristian Šenk

Junij 19, 2017

1. Uvod

Bi rekli, da ima ekipa, ki doseže prvi zadetek, veliko večjo možnost za zmago? Kaj če doseže 4 zadetke? V svetu športa je veliko stvari nepredvidljivih, a iz statističnih podatkov lahko marsikaj tudi predvidimo. Pogledali si bomo podatke o najhitrejšem ekipnem športu na svetu – hokeju. Cilj naloge je, da ugotovimo ali zgornji trditvi res držita. Dodali bomo tudi nekaj izračunov in domnev in tako ugotovili, kako se je hokej spremenil v 100 letih. V začetku se je na tekmah doseglo povprečno 10 golov, danes pa so vratarji dobili veliko večjo vlogo v ekipi, tako se je kljub velikemu napredku v igri in hitrosti povprečno število golov na tekmo skoraj razpolovilo.

Vsi statistični podatki so iz najmočnejše hokejske lige na svetu NHL (National Hockey League) in pridobljeni iz njihove uradne strani nhl.com. Ker so nekateri podatki težko dostopni, sem jih vzel iz naključnega vzorca, ali pa iz zadnje sezone (piše v posamezni nalogi).

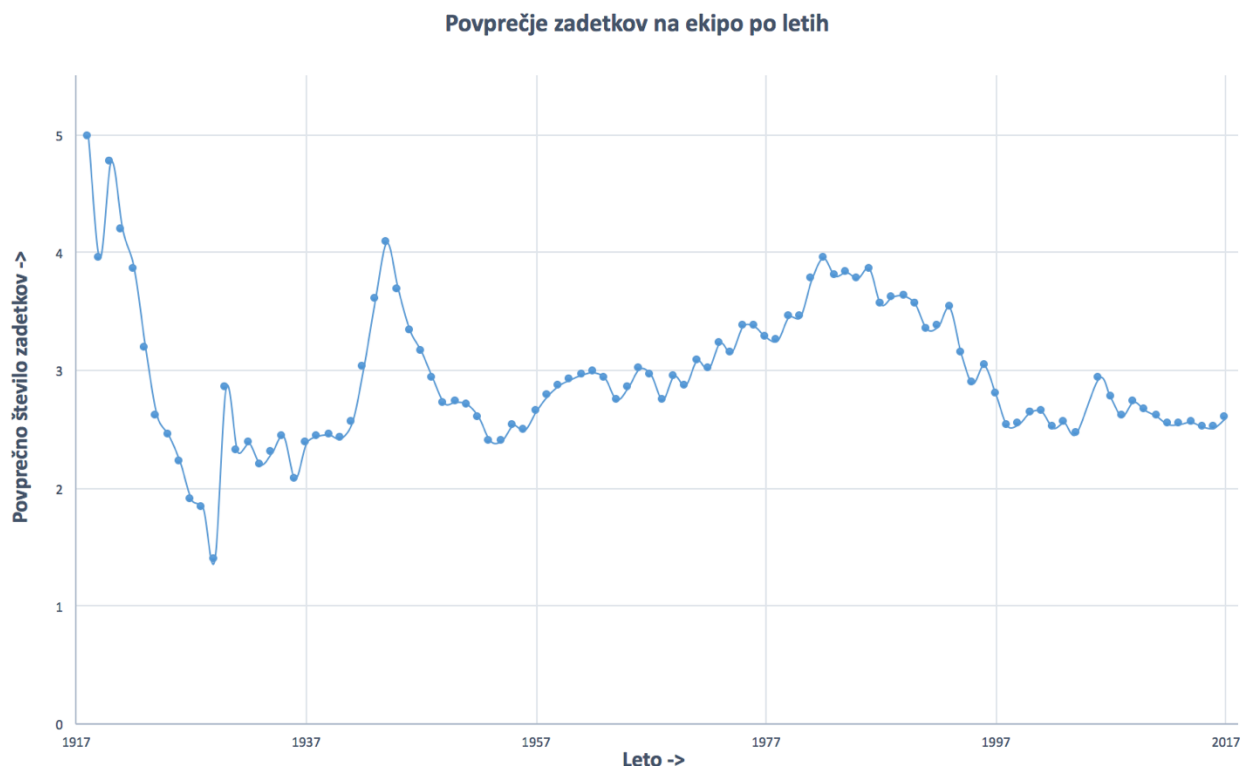
Najprej povejmo nekaj o ligi. Prva sezona NHL-a se je začela leta 1917 in poteka vse do danes. Odigranih je bilo že 54100 tekem. Najprej so sodelovale le štiri ekipe, kasneje se je tekmovanje razširilo na 30 ekip, drugo leto se bo priključila še ekipa iz Vegasa. Tekmovanje poteka v dveh delih. Prvi del je redni del, kjer se ekipe nekajkrat pomerijo vsaka z vsakim, potem se glede na lestvico določijo osmino-finalni pari za končnico (ekipa z največ točkami se bo pomerila z ekipo z najmanj točkami itd.). Zmagovalec je tisti, ki na koncu v finalu zmaga v "best of 7" (prvi doseže štiri zmage) in tako postane zmagovalec Stanlijevega pokala.

2. Statistika

Za začetek si pogledjmo nekaj statističnih podatkov.

Največkrat je bil zmagovalec Stanlijevega pokala (modus) kanadska ekipa Montreal Canadians, ki je od 100ih sezon zmagala kar 25 krat.

V spodnjem grafu si lahko ogledamo, kako se je število zadetkov na ekipo na posamezni tekmi spreminjalo po letih.



Graf 1: povprečje zadetkov na ekipo po letih. Na X-osi imamo podane letnice od prve sezone (1917) do letošnje, na Y-osi je povprečno število zadetkov ekipe na posamezno tekmo. Vidimo, da je bilo maksimalno število zadetkov ravno prvo sezona, torej 1917, saj je ekipa dosegla povprečno 4,98 zadetka na tekmo, najmanj pa leta 1929, ko je bilo povprečje le 1,39.

Omenimo še najpomembnejši statistični parameter – povprečje in njegov standardni odklon. Povprečje zadetkov v ligi je 2,95 zadetka na ekipo, z odklonom 0,6.

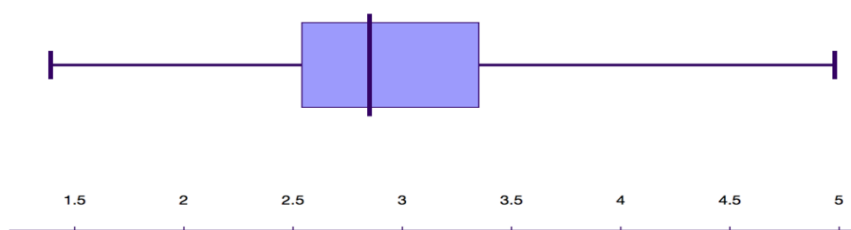
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Poglejmo si še interval zaupanja za povprečje (μ). Za β bomo vzeli 90%.

$$\Delta = \frac{\sigma \cdot t_{(1+\beta)/2}(\infty)}{\sqrt{n}} \quad \bar{x} - \Delta \leq \mu \leq \bar{x} + \Delta$$

$$\Delta = 0,0987 \quad 2,8513 \leq \mu \leq 3,0487$$

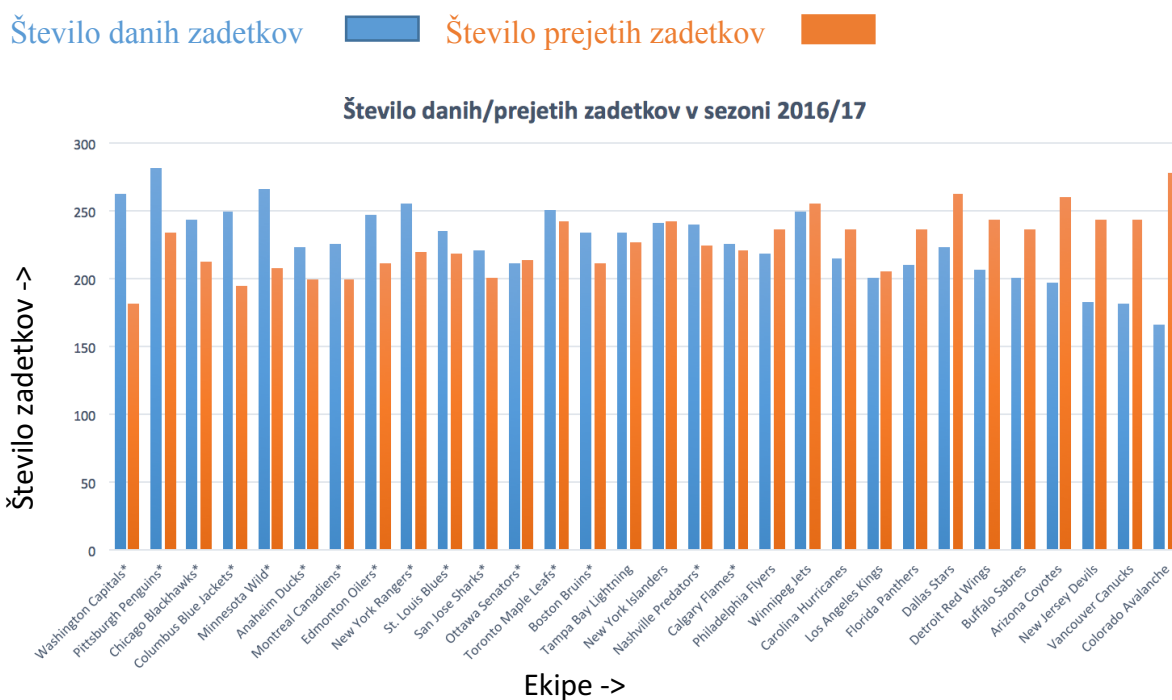
Kvartilni diagram (škatla z brki) izgleda takole:



Povprečno število zadetkov ->

Diagram 1: Kvartilni diagram¹. Prvi kvartil je na 2,5 zadetka, tretji malo pod 3,5.

Spodaj imamo še histogram, ki nam prikazuje, koliko zadetkov je dosegla posamezna ekipa v rednem delu zadnje sezone. Doseženi zadetki so obarvani z modro, prejeti z oranžno barvo. Ekipa so razvrščene glede na razvrstitev po končanem rednem delu tekmovanja.

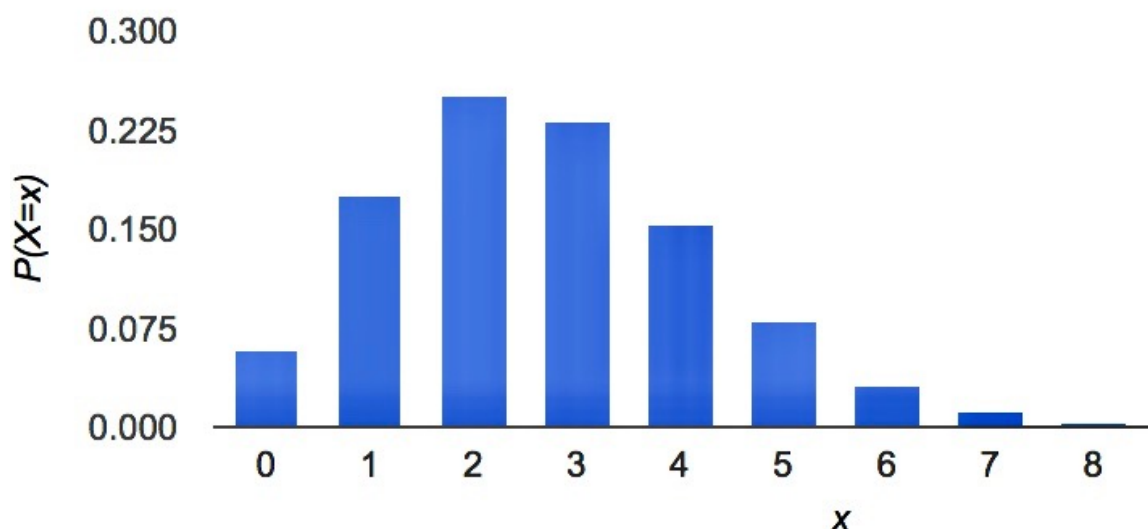


Graf 2: Število danih zadetkov (modra) in prejetih zadetkov (oranžna) v zadnji sezoni

Če bi pogledali samo graf, bi glede na podatke lahko sklepali, da je na prvem mestu pristala ekipa Pittsburgh, na zadnjem pa Colorado Avalanche, a ni tako. Na zadnjem mestu so sicer res pristali igralci Colorada, a prvo mestu po rednem delu so prepričljivo dosegli Capitalsi, ki so sicer na 82tih tekmah dosegli kar 20 zadetkov manj, kot tukaj vodilni Pittsburgh.

¹ Vir [3] stran 124.

Za občutek predpostavimo, da gre za binomsko porazdelitev števila zadetkov. Če za število strel (n) vzamemo 30 in za verjetnost zadetka 9% (oboje je povprečje zadnje sezone).



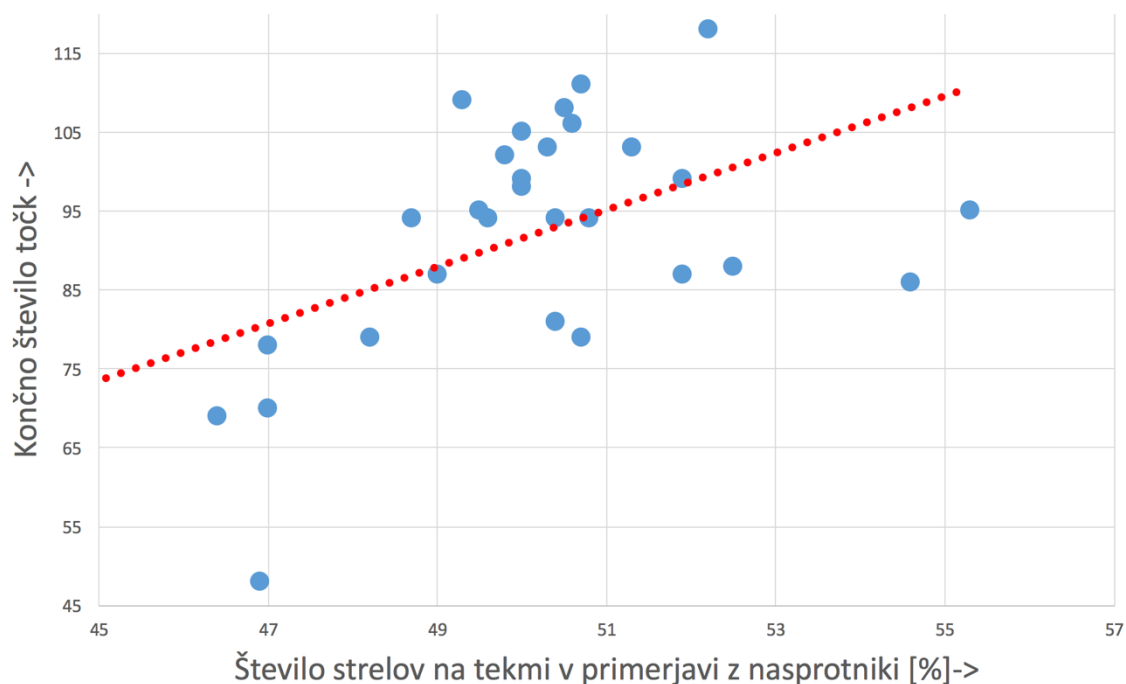
Graf 3: Binomska porazdelitev števila zadetkov na ekipo. Vidimo lahko kako verjetno je, da ekipa doseže določeno število golov na tekmi.

Preden pa izračunamo regresijsko premico med končnim številom točk in številom strel na tekmi, najprej s pomočjo korelacijskega koeficienta izračunajmo, ali sta spremenljivki med seboj sploh odvisni. Če bo rezultat blizu 0, pomeni, da nista odvisna, če pa bo blizu 1 sta.

$$\rho_{XY} = K \frac{(X, Y)}{\sigma_X * \sigma_Y} = \frac{\sum_{i=1}^N (x_i - \mu_X) * (y_i - \mu_Y)}{\sqrt{\sum_{i=1}^N (x_i - \mu_x)^2 * \sum_{i=1}^N (y_i - \mu_Y)^2}}$$

$$\rho_{XY} = 0.534$$

Vidimo, da je rezultat nekje na sredini, ampak se giblje bolj k 1, kar pomeni, da sta spremenljivki do določene mere res povezani.



Graf 4: Na grafu lahko vidimo, kako pomembno je število strel na tekmo. Ekipa, ki strelja več, imajo ponavadi na koncu tudi več točk. Vrednost na X-osi je število strel ekipe na tekmi/število vseh strel na tekmi.

Regresijska premica ima enačbo:

$$\text{Število točk} = 3.6099 \cdot \text{število strel v \%} - 88.991$$

Z enačbo lahko napovemo približno koliko točk bo imela na koncu ekipa, če vemo koliko povprečno streljajo na gol v primerjavi z nasprotniki.

3. Domneve

V vsakem športu je tudi nekaj nenapisanih pravil in nič drugače ni tudi pri hokeju. Vzeli bomo nekaj največkrat uporabljenih in jih s statistiko potrdili ali ovrgli.

1. domneve: Če dosežeš na tekmi vsaj 4 gole je velika verjetnost, da tekmo zmagaš.

Recimo, da za stopnjo zaupanja vzamemo 90% ($1-\alpha$). Če pogledamo celotno statistiko zadnjih 100 let, lahko to domnevo zlahka zavrnemo, saj je bilo večinoma let povprečje zadetkov enako (4), če pa vzamemo podatke le zadnjih nekaj let ugotovimo, da domneva res drži, saj če je ekipa dosegla 4. zadetek je v skoraj 94% zmagala (npr. ekipa, ki je dosegla 4 zadetke je v zadnjem letu od 1240 tekem izgubila le 78 krat) in je zato ne zavrnemo. Če bi število zadetkov povečali na 5, bi se verjetnost še dosti povečala, če bi jo zmanjšali na 3, pa bi ugotovili, da je verjetnost le še malce večja kot pol (skoraj 3 je tudi povprečje danih zadetkov na tekmi).

H_0 ... hipoteza, da je povprečno število zmag vsaj 90%

$$\mu = 0,937$$

$$\delta = 0,24277$$

$$\alpha = 0,10$$

$$TS = ((0,937 - 0,9) / 0,24277) * 35,21 = 5,36656$$

$$K_\alpha = (t_{0,9}(\infty), \infty) = (1,28, \infty)$$

2. domneva: Vratar je več kot pol ekipe.

Za preverjanje te domneve lahko uporabimo več metod. Mi bomo preverjali tako, da bomo pogledali, če so vratarji prvih treh ekip med najboljšimi vratarji v ligi in za zadnje 3 ekipe, če so njihovi vratarji proti koncu lestvice. Če zgornje trditve v večini primerov držijo, bomo to domnevo potrdili, drugače zavrgli.

Vratarji prvih treh ekip v sezoni 2016/17 so se uvrstili na pozicije od 16 do 21, zadnjih na 26, 46 in 64 mestu od 95ih vratarjev. Sicer so vratarji prvih ekip res višje uvrščeni, kot vratarji nižje uvrščenih ekip, a razlika vseeno ni tako velika, da bi hipotezo lahko potrdili. Vzemimo še naključno sezono, npr. 1997/98. Vratarji vodilnih ekip Dallas Stars, New Jersey Devils ter Detroit Red Wings so zasedli 16, 14 ter 21 mesto, vratarji Vancouvra, Floride in Tampe pa 26, 48 ter 46 mesto. Tukaj je bolj opazna razlika med vratarji boljših in slabših ekip, a so vratarji vodilnih ekip še vseeno prenizko na lestvici, da bi lahko domnevo potrdili. Za bolj natančne rezultate sem vzel še dodatnih pet sezon (2011, 1958, 1994, 1951, 1943 - sezone je naključno izbral program v Javi, ki sem ga izdelal za ta primer). V nekaterih primerih so bili najboljši vratarji res v boljše uvrščenih ekipah, a dostikrat so tudi vratarji zadnje uvrščenih ekip nekje na sredini lestvice, zato domnevo na podlagi teh vzorcev zavrnemo.

3. domneva: Ekipa, ki prva zadane, na koncu tudi slavi.

Prvi zadetek na tekmi je vsekakor lepa popotnica za nadaljevanje srečanja, ampak ali je to že dovolj, da bi lahko trdili, da ima potem ta ekipa tudi veliko večjo možnost za zmago? Tokrat bomo nekaj naključnih vzorcev vzeti iz strani http://www.hockey-reference.com/playoffs/WHA_1979.html. Izmed 30 naključnih tekem zadnjih 30ih let –prej ni bilo dovolj popolne statistike (npr. 28.5.2008, 18.5.1988, 18.5.1975, 28.3.1951 so zmagale ekipe, ki so prve zadele, 27.4.1999, 7.5.1993 pa ekipe, ki niso prve zadele) je ekipa, ki je dosegla prvi zadetek zmagala v kar 21 primerih. Zanimivo je tudi dejstvo, da so ekipe do nekje leta 1990, ki so prve zadele, v veliko primerih na koncu tudi slavile, po tem letu pa je prvi zadetek postajal vse manj pomemben. Kljub temu, da ima ekipa, ki prva zadane malce več prednosti, je ta prednost vseeno premajhna, da bi lahko trdili, da bodo skoraj zagotovo zmagali. Prav tako je prvi zadetek vsako leto manj pomemben, zato se ta domneva zavrne.

4. Izračuni

Izračunajmo koliko strellov moramo sprožiti na gol nasprotnika, da bomo z veliko verjetnostjo (90%) zmagali.

Že pri hipotezah smo ugotovili, da so potrebni 4 zadetki, da lahko trdimo, da bomo z veliko verjetnostjo zmagali. Torej sedaj je vprašanje, kolikokrat moramo ustreliti, da bomo dosegli vsaj 4 zadetke. Z uporabo negativne binomske enačbe

$$E(X) = \frac{k}{p}$$

Pri kateri za k vzamemo število zadetkov, za p ne-učinkovitost branjenja posameznega vratarja (100% - povprečna branjenost vratarja) in pridemo do zaključka, da je za najboljše vratarje za to potrebnih skoraj 47 strellov, (Corey Crawford 91,8%), medtem, ko bi za najslabše vratarje v ligi potrebovali zgolj 22 strellov (Malcolm Subban 81,3%). Z uporabo te metode bi lahko zgornjo hipotezo (hipoteza 2) potrdili, namesto ovrgli.

Izračunajmo verjetnost, da ekipa Los Angeles Kings doseže vsaj 2 zadetka v prvi tretini (prvih 20 min)?

Najprej iz podatkov vidimo, da je ekipa LA Kings v celotni sezoni 2016/17 dosegla 201 zadetek, kar je v povprečju 2,45 zadetka na tekmo (82 tekem). Z uporabo Poissonove porazdelitve izračunamo:

$$P(X \geq 2) = 1 - P(X = 1) - P(X = 0) = 1 - \frac{0,82^1}{1!} \cdot e^{-0,82} - \frac{0,82^0}{0!} \cdot e^{-0,82} = 1 - 0,36 - 0,44 = 0,2$$

Če λ vzamemo 2,45/3. Verjetost je torej 20%.

Kolikokrat bi moral streljati najboljši strellec lige (po zadetkih), da bi samo po njegovi zaslugi ekipa skoraj zagotovo slavila?

Najboljši strellec zadnje sezone je Sidney Crosby, ki je dosegel 44 zadetkov od 255 poizkusov, kar pomeni, da je 17,3% uspešen (p). Kolikokrat mora torej streljati, da bi dosegel 4 zadetke (k)? Znova uporabimo negativno binomsko enačbo

$$E(X) = \frac{k}{p}$$

in ugotovimo, da bi moral sprožiti kar 23 strellov, kar je za enega igralca na posamezni tekmi skoraj nemogoče.

Koliko minut povprečno mine med zadetkoma ekipe San Jose Sharks? Na tekmo smo zamudili 10 minut. Kolikšna je verjetnost, da nismo zamudili gola Sharksov?

Sharksi povprečno dajo 2,4 zadetka na tekmo, sepravi naj bi med zadetki bilo nekje 25min. Za izračun drugega vprašanja bomo uporabili eksponentno porazdelitev:

$$P(X > 10) = 1 - P(X \leq 10) = 1 - (1 - e^{-0,04 \cdot 10}) = e^{-0,4} = 0,67$$

Za λ ponovno vzamemo 0,04 (2,4/60). Ugotovimo, da je 2/3 možnosti, da gola Sharksov v prvih 10 minutah ne bomo zamudili.

Prišli smo do prvega kroga končnice, kjer sta se srečala najboljša ekipa rednega dela, ter najslabša. Kolikšna je verjetnost, da napreduje najslabša ekipa, če vemo, da ima le okoli 30% možnosti za zmago? (igra se na best of 7)

Pri tem primeru gre za negativno binomsko porazdelitev. Izračun izgleda takole:

$$P(X = 7) = \binom{7-1}{4-1} \cdot 0,3^4 \cdot (1-0,3)^{7-4} = \binom{6}{3} \cdot 0,3^4 \cdot 0,7^3 = 0,055566$$

Za n smo vzeli 7, za k 4. Najslabša ekipa ima zgolj 5,5% možnosti, da napreduje.

Organizatorji finalnega obračuna za Stanlijev pokal so postavili ceno kart na 75\$ za sedež na balkonu, 150\$ za sedež blizu igrišča, ter 2000\$ za VIP karto. Koliko si lahko obetajo od prodanih kart, če vemo, da je na voljo 4500 sedežev na balkonu, 15 tisoč sedežev blizu igrišča, ter 500 VIP vstopnic? Zapiši porazdelitev in izračunaj kolikšno je matematično upanje?

$$X \sim \begin{pmatrix} 75 & 150 & 2000 \\ 0,225 & 0,75 & 0,025 \end{pmatrix} \quad E(X) = 16,875 + 112,5 + 50 = 179,375$$

Pričakovan zaslužek znaša:

$$E(X) \cdot 20.000 = 3.587.500$$

5. Zaključek

S seminarsko nalogo sem prikazal nekaj statističnih podatkov iz lige NHL (histogram števila zadetkov, graf povprečja števila zadetkov po sezonah, kvartilni diagram, modus, minimum ter maksimum števila zadetkov), preveril sem nekaj domnev, ter izračunal verjetnosti za določene dogodke, ki se pojavljajo ob igranju hokeja. Ugotovil sem, da določene domneve niso potrjene s strani statističnih podatkov, nekatere pa veljajo. Pri izračunih verjetnosti sem ugotovil, kako malo verjetni so nekateri dogodki, ki se lahko zgodijo na tekmah.

Viri

- [1] nhl.com (19.7.2017)
(<http://www.nhl.com/stats/team?aggregate=0&gameType=3&report=teamssummary&reportType=season&seasonFrom=20162017&seasonTo=20162017&filter=gamesPlayed,gte,&sort=points,wins>) (19.7.2017)
- [2] hockey-references.com (<http://www.hockey-reference.com/leagues/>) (19.7.2017)
- [3] Aleksandar Jurišić, Verjetnostni račun in statistika, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, 2015
- [4] David S. Moore, Purdue University, Statistika: Znanost o podatkih