

Sveučilište u Zagrebu  
Prirodoslovno-matematički fakultet  
Matematika

Statistika

# **Raznolika analiza LPL Spring Split 2021**

Tin Benaković, Jan Corazza, Ante Čubela, Ivan Glavaš

Lipanj 2021.

## Sadržaj

1 Uvod .....	1
2 Prikupljanje podataka .....	3
3 Opisna statistika .....	4
4 Linearna regresija .....	11
5 Statistički testovi .....	16
6 Tablica podataka.....	19
7 Izvori.....	20

# 1 Uvod

League of Legends (nadalje LoL) je kompetitivna mrežna video igra s najvećim međunarodnim e-sport turnirima. Na turnirima se natječu timovi od 5 igrača u mečevima koji u prosjeku traju pola sata. Svaki meč igraju dva tima i pobjednik je uvijek poznat. Uvjet pobjede je uništenje neprijateljskog *Nexusa* tj. baze, a kako bi tim došao do baze potrebno je prethodno uništiti određen broj protivničkih kula kroz jednu od tri staze. Na putu do pobjede timovi se također natječu u osvajanju podciljeva koji se otključavaju nakon određenog vremena i donose prednosti timu koji ih ostvari (Zmaj, Baron, Herald). Svaki igrač kontrolira unaprijed odabranog "heroja" s popisa i timovi se tokom igre koordiniraju glasovno.



Odлучili smo promatrati sljedeće varijable: postotak pobjeda(WR), postotak uzimanja prve kule (FT), postotak uzimanja prve tri kule (F3T),

postotak uzimanja prvog Zmaja (FD), postotak uzimanja *Rift Herald*-a (HLD), omjer obaranja i smrti (KD), prosječno vrijeme trajanja meča (AGT) i postotak uzimanja *Baron Nashor*-a (BN), jer ih smatramo najrelevantnijim opisima učinka tokom turnira.

Neke od prikupljenih podataka podvrgnuli smo analizi metodama iz opisne statistike. Uradili smo Pearsonov test koreliranosti za sljedeće parove podataka: postotak uzimanja prvog Zmaja/postotak uzimanja *Rift Herald*-a, postotak uzimanja prve tri kule/omjer obaranja i smrti, postotak pobjeda/prosječno trajanje meča. Za sve navedene parove podataka smo pokušali predvidjeti Pearsonov koeficijent korelacije i naći objašnjenje naših rezultata.

Napravili smo linearnu regresiju za podatke o postotku uzimanja prve tri kule i omjeru uzimanja i smrti, te za podatke o postotku uzimanja prvog Zmaja i postotku uzimanja *Rift Herald*-a.

Također smo obavili određene statističke testove ( $t$ -test,  $F$ -test,  $KS$ -test) za varijable za koje su ti testovi prikladni.

## 2 Prikupljanje podataka

Prikupili smo podatke o timovima Kineske profesionalne LoL lige u sezoni LPL Spring Split 2021. Tih 17 timova koji su sudjelovali u ligi te sezone će predstavljati uzorak našeg istraživanja. Potrebne podatke smo prikupili iz službene baze podataka odigranih mečeva koju se može pronaći na poveznici na kraju seminara.

Svaki tim, tj. uzorak iz populacije, predstavljen je kao slučajni vektor s određenim brojem komponenti koje opisuju učinak tog tima na turniru. Varijable na koje se referiramo su stupci u matičnom zapisu podataka danom na kraju seminara. Relevantne stupce koje smo koristili izvadili smo u tablici koja se nalazi na kraju seminara.

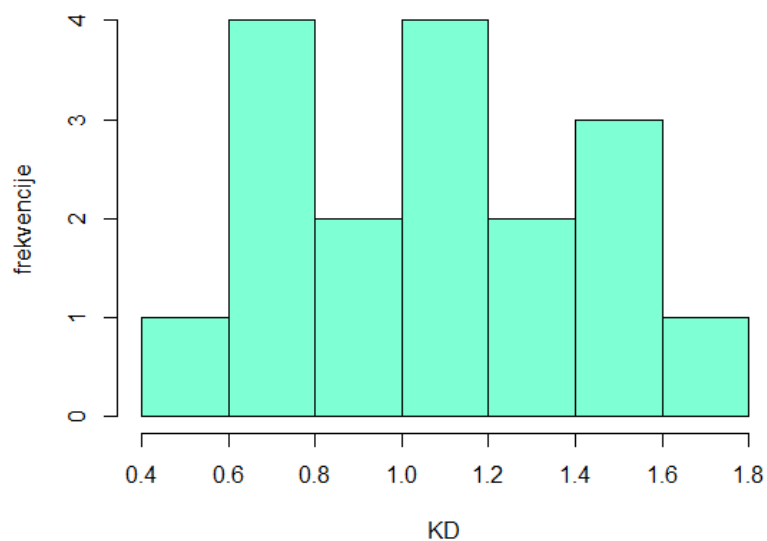
### 3 Opisna statistika

Promatramo varijablu KD koja svakom timu pridružuje broj koji se dobije tako da ukupan broj obaranja neprijatelja podijelimo s ukupnim brojem smrti toga tima kroz sezonu. Vrijednost te varijable je neki (racionalan) nenegativan broj. Ukoliko je broj smrti tima tokom sezone 0, kao nazivnik broja KD se uzima 1.

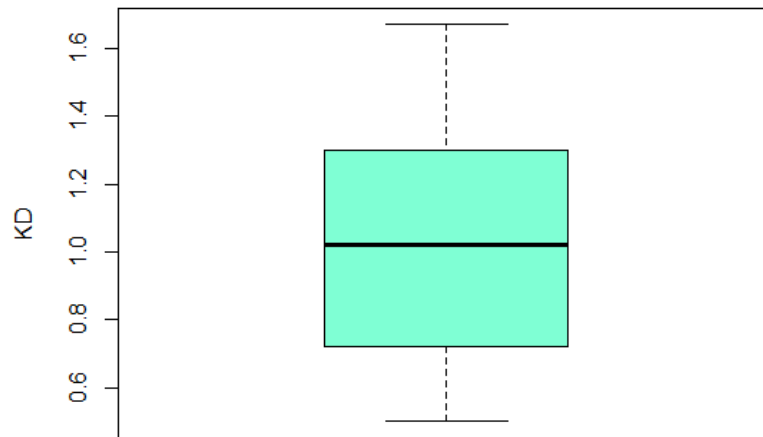
Karakteristična petorka varijable KD je:

$$(0.500, 0.715, 1.020, 1.375, 1.670) .$$

Histogram i dijagram pravokutnika varijable KD su dani ispod:



Slika 1: *Histogram varijable KD*

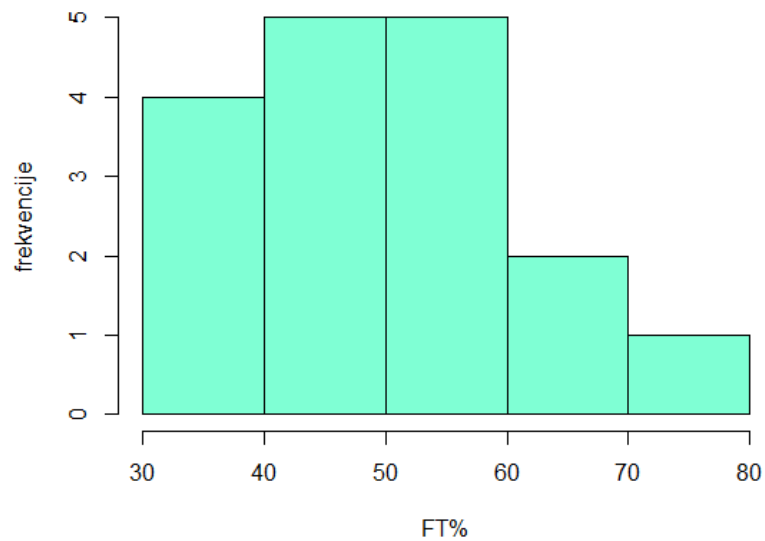


Slika 2: *Dijagram pravokutnika varijable KD*

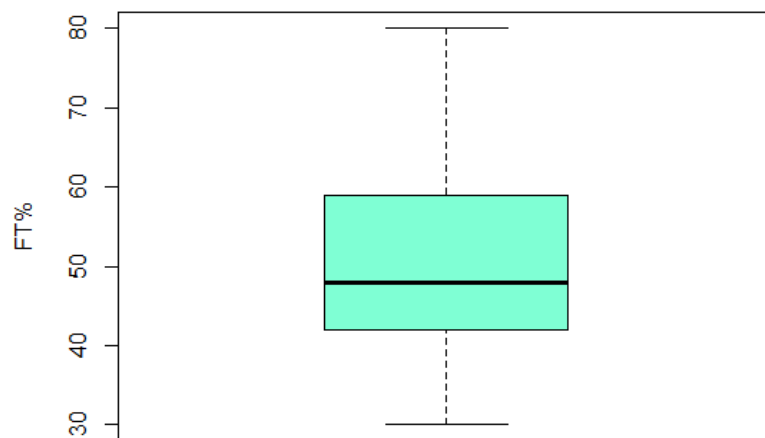
Promatramo varijablu FT koja svakom timu pridružuje broj koji je jednak postotku mečeva u kojima je taj tim srušio prvu kulu. Karakteristična petorka varijable FT je:

$$(30.0, 39.0, 48.0, 59.5, 80.0) .$$

Histogram i dijagram pravokutnika varijable FT su dani ispod:



Slika 3: *Histogram varijable FT*



Slika 4: *Dijagram pravokutnika varijable FT*

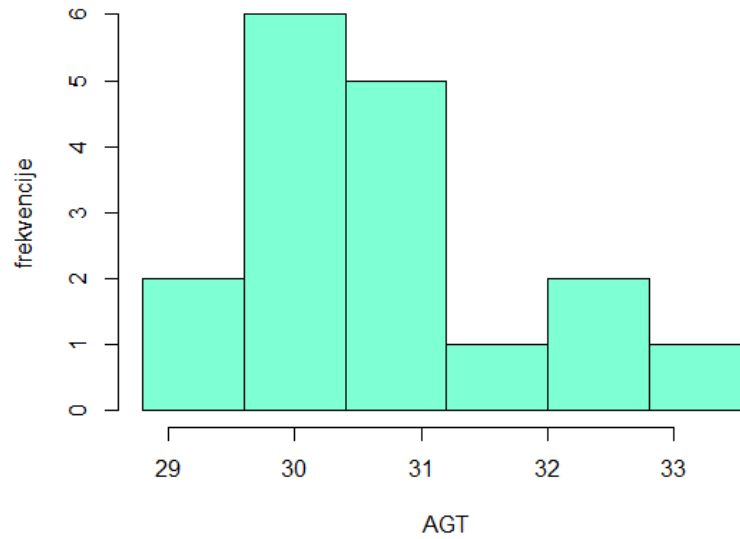
Promatramo varijablu AGT koja svakom timu pridružuje broj pobjeda koje je tim ostvario tijekom sezone. Karakteristična petorka varijable



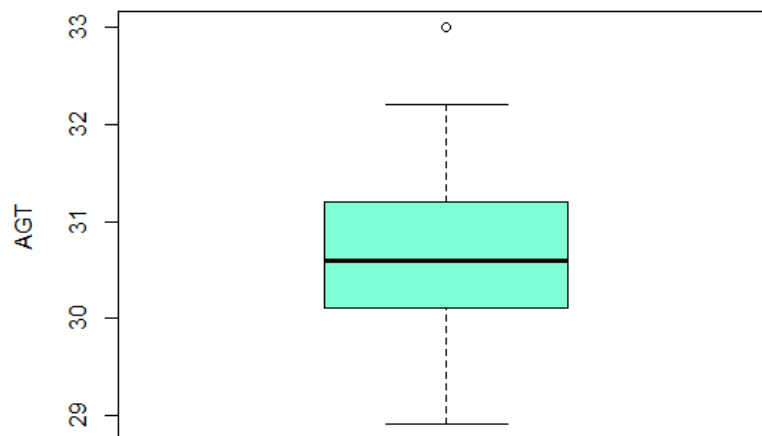
AGT je:

$(28.90, 30.05, 30.60, 31.50, 33.00)$ .

Histogram i dijagram pravokutnika varijable AGT su dani ispod:



Slika 5: *Histogram varijable AGT*



Slika 6: *Dijagram pravokutnika varijable AGT*

Napravili smo Pearsonov test korelacije za sljedeće parove varijabli:

- FT / KD
- FD / HLD
- WR / AGT

Pearsonov koeficijent korelacije se definira na sljedeći način:

$$r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}},$$

gdje je:

$$\begin{aligned} S_{XX} &= \sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \\ S_{YY} &= \sum_{i=1}^n y_i^2 - n \cdot \bar{y}^2 \\ S_{XY} &= \sum_{i=1}^n x_i y_i - n \cdot \bar{x} \bar{y}. \end{aligned}$$

U našem slučaju,  $n = 17$  je broj timova, a  $x_i$  i  $y_i$  su mjerenja prve odnosno druge varijable čiju koreliranost mjerimo.

Jedna varijabla (F3T) za koju se smatra da dobro prati efektivnost tima kroz igre je postotak igara u kojima su protivničkom timu uništili tri kule prije nego što je protivnički tim to napravio njima (razlog je to da je uništavanje kula nužan korak prema osvajanju meča). Još jedna takva varijabla (KD) je poznati *KD ratio* tj. omjer obaranja protivničkih heroja naspram smrti heroja iz tima. Očekivali smo da će te varijable biti pozitivno korelirane; dobri timovi će u većem udjelu igara biti prvi koji će uništiti tri protivničke kugle i također će imati više obaranja i manje vlastitih smrti.

Izračunali smo Pearsonov koeficijent za variable F3T i KD:

$$r \approx 0.7824.$$

Čime smo potvrdili našu pretpostavku.

Varijable FD i HLD označavaju postotak mečeva u kojima je tim osvojio prvog Zmaja (FD, *First Drake*) i postotak osvojenih *Rift Herald*-a respektivno.

Zmaj pojačava sve igrače u jednom timu, a Herald se koristi kako bi se rušile tuđe kule. Ta dva cilja se najčešće osvajaju u isto vrijeme, ali na suprotnim stranama mape. Zbog toga pretpostavljamo da ako jedan tim se više fokusira na to da uzme prvog zmaja će morati prepustiti protivničkom timu da uzmu Heralda i da ćemo dobiti negativnu korelaciju između te dvije varijable. Očekujemo da apsolutna vrijednost korelacije nije jako velika jer će bolji timovi često uspjevati i osigurati oba cilja za sebe.

Izračunali smo Pearsonov koeficijent za varijable FD i HLD:

$$r \approx -0.2630.$$

Što potvrđuje našu pretpostavku.

Varijabla WR označava postotak pobjeda određenog tima, a varijabla AGT prosječno trajanje meča određenog tima. Teško je točno predvidjeti hoće li te dvije varijable biti povezane, ali pretpostavljamo da neće biti jako povezane. Razlog tome je što će najbolji timovi u ligi jako brzo pobjeđivati svoje mečeve, ali će i najslabiji timovi u ligi jako brzo gubiti svoje mečeve. Očekuje se da je duljina trajanja meča malo veća kod timova koji su po jakosti jako bliski ili možda prema postotku pobjeda

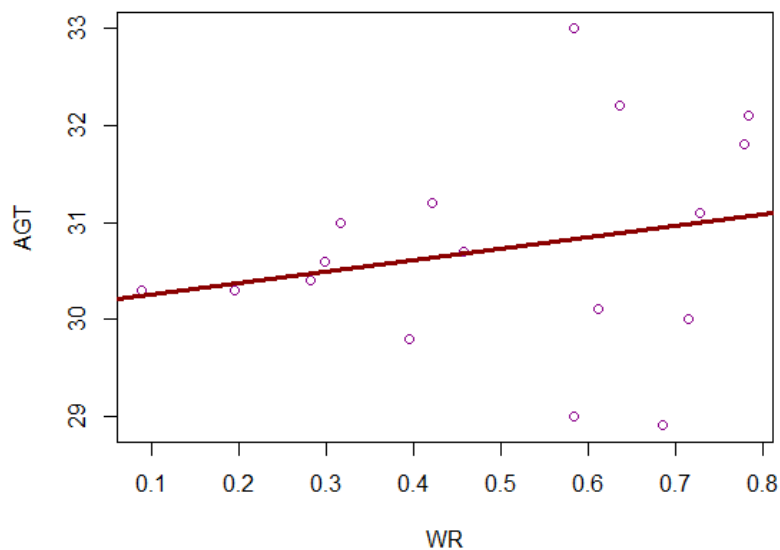
blizu medijana varijable WR.

Izračunali smo Pearsonov koeficijent za variable WR i AGT:

$$r \approx 0.2293.$$

Što potvrđuje našu pretpostavku.

Slika podataka WR i AGT prikazanih u koordinatnom sustavu, i pravca koji aproksimira te podatke, je dana ispod:



Slika 7: *Linearna regresija varijabli WR i AGT*

## 4 Linearna regresija

Procijenili smo parametre linearne regresije i nacrtali regresijski pravac za varijable F3T i KD. Podaci o spomenutim varijablama za timove su:

<b>F3T</b>	47	67	19	83	56	61	30	49
<b>KD</b>	0.81	1.67	0.67	1.30	1.16	1.23	0.72	1.01
28	53	12	59	42	56	45	89	58
0.68	1.02	0.50	1.54	1.45	1.18	0.71	1.58	0.85

Neka je linearni model koji aproksimira dane podatke dan s:

$$y = \alpha + \beta x + \varepsilon,$$

gdje  $y$  predstavlja varijablu KD, a  $x$  predstavlja varijablu F3T. Koeficijenti  $\alpha$  i  $\beta$  su parametri modela, a  $\varepsilon$  je slučajna pogreška. Najprije smo odredili sljedeće podatke koji su nam potrebni:

$$\bar{x} = 50.23529$$

$$\bar{y} = 1.063529$$

$$S_{xx} = 6633.059$$

$$S_{yy} = 2.094588$$

$$S_{xy} = 92.22588.$$

Parametre linearne regresije smo dobili na sljedeći način:

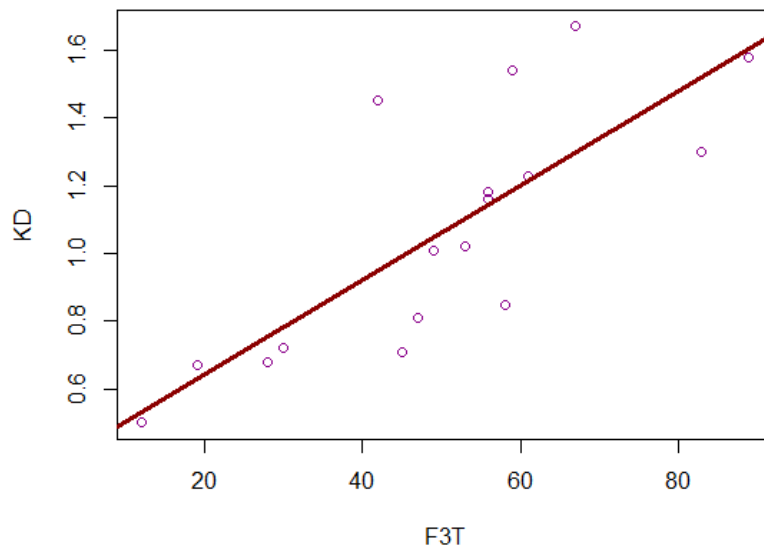
$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = 0.01390397$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 0.3650592.$$

Slijedi da je jednačba regresijskog pravca sljedeća:

$$y = \hat{\alpha} + \hat{\beta}x = 0.3650592 + 0.01390397x.$$

Regresijski pravac s nacrtanim podacima ( $F3T, KD$ ) za svaki od 17 timova je dan na slici ispod.



Slika 8: Linearna regresija varijabli  $F3T$  i  $KD$

Odredili smo parametar  $\hat{\sigma}$  na sljedeći način:

$$SSE = Syy - \hat{\beta}^2 Sxx = 0.8122819$$

$$\hat{\sigma} = \sqrt{\frac{SSE}{n-2}} = 0.2327061.$$

Uz takav parametar  $\sigma$  znamo da vrijedi:

$$\frac{\hat{\alpha} - \alpha}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim t(n-2), \quad \frac{\hat{\beta} - \beta}{\hat{\sigma} \sqrt{\frac{1}{S_{xx}}}} \sim t(n-2).$$

Za pouzdanost  $\alpha = 0.05$  i  $n = 17$ , imamo  $t_{\frac{\alpha}{2}}(n-2) \approx 2.1314$ , pa znamo:

$$\mathbb{P} \left( -2.1314 \leq \frac{\hat{\alpha} - \alpha}{\hat{\sigma} \sqrt{\frac{1}{17} + \frac{\bar{x}^2}{S_{xx}}}} \leq 2.1314 \right) = 0.95,$$

to jest:

$$\mathbb{P} \left( \hat{\alpha} - 2.1314\hat{\sigma} \sqrt{\frac{1}{17} + \frac{\bar{x}^2}{S_{xx}}} \leq \alpha \leq \hat{\alpha} + 2.1314\hat{\sigma} \sqrt{\frac{1}{17} + \frac{\bar{x}^2}{S_{xx}}} \right) = 0.95.$$

Slijedi da je 95%-tni interval pouzdanosti za parametar  $\alpha$ :

$$[0.03633, 0.69379].$$

Analognim postupkom dobijemo:

$$\mathbb{P} \left( \hat{\beta} - 2.1314\hat{\sigma} \sqrt{\frac{1}{S_{xx}}} \leq \beta \leq \hat{\beta} + 2.1314\hat{\sigma} \sqrt{\frac{1}{S_{xx}}} \right) = 0.95.$$

Slijedi da je 95%-tni interval pouzdanosti za parametar  $\beta$ :

$$[0.00781, 0.01999].$$

Uz pretpostavku da  $\varepsilon$  ima  $N(0, \sigma^2)$  distribuciju, vrijedi:

$$(n-2) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2),$$

pa smo kao procjenu 90%-tnog intervala pouzdanosti za  $\sigma^2$  dobili:

$$\mathbb{P} \left( 7.2609 \leq 15 \cdot \frac{\hat{\sigma}^2}{\sigma^2} \leq 24.9958 \right) = 0.9,$$

to jest:

$$\mathbb{P} \left( \frac{15\hat{\sigma}^2}{24.9958} \leq \sigma^2 \leq \frac{15\hat{\sigma}^2}{7.2609} \right) = 0.9.$$

Slijedi da je 90%-tni interval za  $\sigma^2$ :

$$[0.03250, 0.11187].$$

Proveli smo i test značajnosti ovoga linearnog regresijskog modela, na nivou značajnosti 0.05. Naše hipoteze su:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0.$$

Prethodno smo dobili da je 95% pouzdani interval za parametar  $\beta$  interval  $[0.00781, 0.01999]$ . Kako  $0 \notin [0.00781, 0.01999]$ , zaključili smo da možemo odbaciti hipotezu  $H_0$  u korist hipoteze  $H_1$ , na nivou značajnosti 0.05.

Procijenili smo parametre linearne regresije i nacrtali regresijski pravac za varijable FD i HLD. Podaci o spomenutim varijablama za timove su:

<b>FD</b>	47	61	39	37	72	45	62	40
<b>HLD</b>	32	54	42	60	43	47	37	47
56	53	41	51	48	53	55	49	37
45	45	52	49	57	58	57	67	60

Analognim postupkom kao za prethodni primjer, dobili smo da je jednadžba regresijskog pravca sljedeća:

$$y = \hat{\alpha} + \hat{\beta}x = 62.48501 - 0.2485168x.$$

Kao 90%-tne intervale pouzdanosti za parametre  $\alpha$ ,  $\beta$  i  $\sigma^2$  dobijemo:

$$\alpha \in [41.58357, 83.38644049]$$

$$\beta \in [-0.66120, 0.16417]$$

$$\sigma^2 \in [50.11593, 172.2879].$$

Također smo proveli i test značajnosti ovoga linearnog regresijskog modela, na nivou značajnosti 0.1. Naše hipoteze su:

$$H_0 : \beta = 0$$

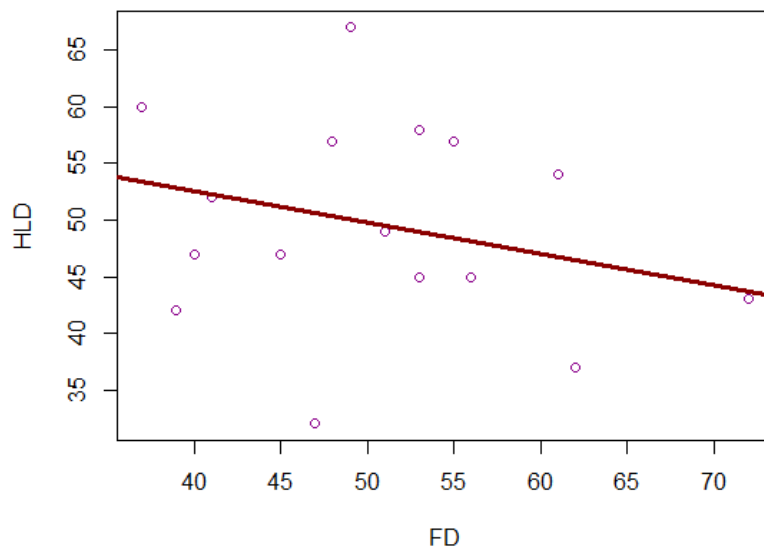
$$H_1 : \beta \neq 0.$$

Prethodno smo dobili da je 90% pouzdani interval za parametar  $\beta$  interval  $[-0.66120, 0.16417]$ . Kako  $0 \in [-0.66120, 0.16417]$ , zaključili smo da ne



možemo odbaciti hipotezu  $H_0$  u korist hipoteze  $H_1$ , na nivou značajnosti 0.1.

Regresijski pravac sa nacrtanim podacima  $(FD, HLD)$  za svaki od 17 timova je dan na slici ispod.



Slika 9: *Linearna regresija varijabli FD i HLD*

## 5 Statistički testovi

Unutar meča postoje podciljevi čije osvajanje olakšava pobjedu. Jedan takav podcilj je ubijanje zmaja. Varijabla koja je vezana uz osvajanje tog cilja je FD tj. postotak igara u kojima je tim prvi ubio zmaja. Pošto je postizanje tog cilja nešto što očekujemo da bude ostvareno u većini mečeva na kompetetivnoj razini u koju spada ova liga, prirodno je zapitati se je li očekivanje FD upravo 50% (zato što postizanje jednog tima povlači nepostizanje protivničkog). Za ispitivanje te pretpostavke koristili smo dvostrani t-test.

$$H_0 : E[FD] = 50\%$$

$$H_1 : E[FD] \neq 50\%.$$

Proveli smo prikladni test u R-u na razini pouzdanosti 95% tj.  $\alpha = 0.05$  i dobili sljedeće rezultate:

- p-vrijednost = 0.92
- 95% interval pouzdanosti: [44.77, 54.75],

koji indiciraju da ne možemo odbaciti nultu hipotezu da očekivanje od FD nije 50%.

Slično kao i za varijablu FD pretpostavljamo da će i varijabla BN (udio u kontroli Barona, još jedan podcilj kojeg na kompetetivnoj razini timovi ostvaruju) imati očekivanje 50%, iz sličnih razloga kao i varijabla FD. Zanimljivo je ustanoviti odnos varijanca tih varijabli i kako bi to ispitali smo koristili F-test. Hipoteze našeg testa su:

$$H_0 : Var[FD] = Var[BN]$$

$$H_1 : Var[FD] \neq Var[BN].$$

Rezultantna p-vrijednost je 0.03394 i shodno tome odbacujemo nul-hipotezu u korist alternativne i zaključujemo da su varijance ovih varijabli različite s barem 95% pouzdanosti. To je smislen zaključak i možemo ga objasniti kao posljedicu pravila igre: naime, Baron se pojavljuje u 20. minuti igre, dok se Zmaj pojavljuje u ranoj igri. Timovi se u tih 20 minuta igre opremljuju i napreduju, i tim sastavljen od boljih igrača ima priliku znatno uvećati svoju apriornu prednost. Varijanca FD-a je relativno malena jer se timovi nisu stigli opremiti i događaj više nasumičan (dakle koncentriran oko 50%) od onog kasnijeg gdje će dobri timovi najčešće imati veću vrijednost a lošiji najčešće malenu (dakle sveukupne vrijednosti će biti veoma raspršene oko očekivanog 50%). Također smo napravili jednostrane F-testove na te dvije varijable da bi saznali nešto više o relativnim veličinama:

$$H_0 : Var[FD] = Var[BN]$$

$$H_1 : Var[FD] > Var[BN].$$

Za ovu alternativnu hipotezu dobili smo p-vrijednost 0.983.

$$H_0 : Var[FD] = Var[BN]$$

$$H_1 : Var[FD] < Var[BN].$$

Za ovu alternativnu hipotezu dobili smo p-vrijednost 0.01697. Oba su rezultata konzistentna s gornjim objašnjenjem.

Odradili smo Kolmogorov-Smirnovljev test da provjerimo ima li varijabla WPM (*wards-per-minute*) normalnu razdiobu. Uradili smo to tako da smo standardizirali varijablu WPM i provjerili ima li novo dobivena varijabla jediničnu normalnu razdiobu. Neka je  $F$  funkcija gustoće te

ново dobivene varijable. Hipoteze našeg testa su:

$$H_0 : F = F_0$$

$$H_1 : \text{ne } H_0,$$

gdje je:

$$F_0(x) = \Phi(x),$$

to jest novo dobivena varijabla ima jediničnu normalnu razdiobu. Kao rezultat testa smo dobili:

$$p = 0.4229,$$

pa smo zaključili da ne možemo odbaciti hipotezu  $H_0$ , to jest da ne možemo odbaciti mogućnost da WPM ima normalnu razdiobu.

## 6 Tablica podataka

Ime tima	WR	KD	FT	F3T	AGT	FD	HLD	BN
Bilibili Gaming	0.421	0.81	45%	47%	31.2	47%	32%	48%
Edward Gaming	0.778	1.67	61%	67%	31.8	61%	54%	60%
eStar	0.194	0.67	33%	19%	30.3	39%	42%	15%
FunPlus Phoenix	0.686	1.30	66%	83%	28.9	37%	60%	68%
Invictus Gaming	0.583	1.16	44%	56%	29.0	72%	43%	68%
JD Gaming	0.727	1.23	48%	61%	31.1	45%	47%	51%
LGD Gaming	0.297	0.72	30%	30%	30.6	62%	37%	31%
LNG Esports	0.457	1.01	60%	49%	30.7	40%	47%	37%
Oh My God	0.282	0.68	36%	28%	30.4	56%	45%	43%
Rare Atom	0.583	1.02	47%	53%	33.0	53%	45%	57%
Rogue Warriors	0.088	0.50	32%	12%	30.3	41%	52%	21%
Royal Never Give Up	0.784	1.54	59%	59%	32.1	51%	49%	75%
Suning	0.636	1.45	42%	42%	32.2	48%	57%	57%
Team WE	0.611	1.18	58%	56%	30.1	53%	58%	59%
ThunderTalk Gaming	0.316	0.71	53%	45%	31.0	55%	57%	45%
Top Esports	0.714	1.58	80%	89%	30.0	49%	67%	67%
Victory Five	0.395	0.85	55%	58%	29.8	37%	60%	45%

## 7 Izvori

(1) Kolegij Statistika

<https://web.math.pmf.unizg.hr/nastava/stat/>

(2) LoL Esports

<https://lolesports.com/>

(3) R dokumentacija

<https://www.r-project.org/other-docs.html>