

Statistička analiza podataka - projekt

Sapunanje

2022-12-18

Statistika nogometaša engleske Premier lige

Studenti: Karlo Boroš, Petar Novak, Vlado Perković i Mislav Rendulić

1. Uvod

Ovaj projekt iz kolegija Statistička analiza podataka radili smo pod vodstvom asistenta *insert_name*. Ovaj seminar ćemo podijeliti u par dijelova:

1. Uvod
2. Osnovna prilagodba podataka
3. Generalne informacije o podacima
4. XY-test
5. YZ-test
6. ZX-test
7. XZ-test
8. Rezultati

Imali smo sreće i dobili smo upravo zadatak koji smo i priželjkivali.

Cilj ovoga projekta je uzeti dane podatke i iz njih probati izvući zaključke i faktore koji mogu utjecati na rezultat, broj golova i sl. Naravno, nije potrebno naglasiti važnost korištenja ispravnih testova te dobivanje rezultata koji su validni.

2. Osnovna prilagodba podataka

Podatke je prvo potrebno učitati. Bitno je dobro ih proučiti kako ne bismo slučajno pogriješili u nekom zaključku. Nakon dobre analize možemo krenuti sa našim zadacima.

izbrisati kasnije *napomena ostatku ekipe: spremio sam podatke kao dataset.csv jer je ime dugo i ne očitava š pa je ovo najjednostavnije*

#poslije dodati include = FALSE, za sad nek ostane da vidimo sve sta treba

```
nogometasi <- read.csv('dataset.csv', encoding = "UTF-8")
nogometasi$Nation <- str_sub(nogometasi$Nation, -3)
head(nogometasi)
```

##		Player	Team	Nation	Pos	Age	MP	Starts	Min	X90s	Gls	Ast	G.PK		
## 1		Bukayo Saka	Arsenal	ENG	FW,MF	19	38	36	2,978	33.1	11	7	9		
## 2		Gabriel Dos Santos	Arsenal	BRA	DF	23	35	35	3,063	34.0	5	0	5		
## 3		Aaron Ramsdale	Arsenal	ENG	GK	23	34	34	3,060	34.0	0	0	0		
## 4		Ben White	Arsenal	ENG	DF	23	32	32	2,880	32.0	0	0	0		
## 5		Martin Ødegaard	Arsenal	NOR	MF	22	36	32	2,785	30.9	7	4	7		
## 6		Granit Khaka	Arsenal	SUI	MF,DF	28	27	27	2,327	25.9	1	2	1		
##		PK	PKatt	CrdrY	CrdrR	Gls.1	Ast.1	G.A	G.PK.1	G.A.PK	xG	npxG	xA	npxG.xA	xG.1
## 1	2	2	6	0	0.33	0.21	0.54	0.27	0.48	9.7	8.2	6.9	15.2	0.29	

```
## 2 0 0 8 1 0.15 0.00 0.15 0.15 0.15 2.7 2.7 0.8 3.5 0.08
## 3 0 0 1 0 0.00 0.00 0.00 0.00 0.00 0.0 0.0 0.0 0.0 0.00
## 4 0 0 3 0 0.00 0.00 0.00 0.00 0.00 1.0 1.0 0.6 1.6 0.03
## 5 0 0 4 0 0.23 0.13 0.36 0.23 0.36 4.8 4.8 6.8 11.6 0.16
## 6 0 0 10 1 0.04 0.08 0.12 0.04 0.12 1.2 1.2 2.3 3.5 0.05
## xA.1 xG.xA npxG.1 npxG.xA.1
## 1 0.21 0.50 0.25 0.46
## 2 0.02 0.10 0.08 0.10
## 3 0.00 0.00 0.00 0.00
## 4 0.02 0.05 0.03 0.05
## 5 0.22 0.38 0.16 0.38
## 6 0.09 0.14 0.05 0.14
```

```
str(nogometasi)
```

```
## 'data.frame': 691 obs. of 30 variables:
## $ Player : chr "Bukayo Saka" "Gabriel Dos Santos" "Aaron Ramsdale" "Ben White" ...
## $ Team : chr "Arsenal" "Arsenal" "Arsenal" "Arsenal" ...
## $ Nation : chr "ENG" "BRA" "ENG" "ENG" ...
## $ Pos : chr "FW,MF" "DF" "GK" "DF" ...
## $ Age : int 19 23 23 23 22 28 28 24 21 20 ...
## $ MP : int 38 35 34 32 36 27 24 22 33 29 ...
## $ Starts : int 36 35 34 32 32 27 23 22 21 21 ...
## $ Min : chr "2,978" "3,063" "3,060" "2,880" ...
## $ X90s : num 33.1 34 34 32 30.9 25.9 22.5 21.3 21.3 20.7 ...
## $ GlS : int 11 5 0 0 7 1 2 1 10 6 ...
## $ Ast : int 7 0 0 0 4 2 1 3 2 6 ...
## $ G.PK : int 9 5 0 0 7 1 2 1 10 5 ...
## $ PK : int 2 0 0 0 0 0 0 0 0 1 ...
## $ PKatt : int 2 0 0 0 0 0 0 0 0 1 ...
## $ CrdY : int 6 8 1 3 4 10 6 0 1 3 ...
## $ CrdR : int 0 1 0 0 0 1 0 0 0 1 ...
## $ GlS.1 : num 0.33 0.15 0 0 0.23 0.04 0.09 0.05 0.47 0.29 ...
## $ Ast.1 : num 0.21 0 0 0 0.13 0.08 0.04 0.14 0.09 0.29 ...
## $ G.A : num 0.54 0.15 0 0 0.36 0.12 0.13 0.19 0.56 0.58 ...
## $ G.PK.1 : num 0.27 0.15 0 0 0.23 0.04 0.09 0.05 0.47 0.24 ...
## $ G.A.PK : num 0.48 0.15 0 0 0.36 0.12 0.13 0.19 0.56 0.53 ...
## $ xG : num 9.7 2.7 0 1 4.8 1.2 2.5 0.7 5.8 7.2 ...
## $ npxG : num 8.2 2.7 0 1 4.8 1.2 2.5 0.7 5.8 6.5 ...
## $ xA : num 6.9 0.8 0 0.6 6.8 2.3 1.3 1.9 2.2 3.3 ...
## $ npxG.xA : num 15.2 3.5 0 1.6 11.6 3.5 3.8 2.6 8 9.8 ...
## $ xG.1 : num 0.29 0.08 0 0.03 0.16 0.05 0.11 0.03 0.27 0.35 ...
## $ xA.1 : num 0.21 0.02 0 0.02 0.22 0.09 0.06 0.09 0.1 0.16 ...
## $ xG.xA : num 0.5 0.1 0 0.05 0.38 0.14 0.17 0.12 0.37 0.51 ...
## $ npxG.1 : num 0.25 0.08 0 0.03 0.16 0.05 0.11 0.03 0.27 0.31 ...
## $ npxG.xA.1 : num 0.46 0.1 0 0.05 0.38 0.14 0.17 0.12 0.37 0.47 ...
```

Nakon enkodiranja početnih podataka, postajala su odstupanja od stvarnih imena kod nekih igrača pa smo ta imena ručno ispravili.

```
#utf-8 encoding
```

```
#imena <- c("Halil Dervišoğlu", "Michał Karbownik", "Matěj Vydra", "Mateo Kovačić", "Luka Milivojević")
```

```
#ako nema elegantnijeg nacina nek ovo ostane
```

```

nogometasi$Player[nogometasi$Player == "Halil Dervi?o?lu"] <- "Halil Dervi?o?lu"
nogometasi$Player[nogometasi$Player == "?lkay Gündo?an"] <- "İlkay Gündoğan"
nogometasi$Player[nogometasi$Player == "?ukasz Fabia?ski"] <- "Łukasz Fabiański"
nogometasi$Player[nogometasi$Player == "Asmir Begovi?"] <- "Asmir Begović"
nogometasi$Player[nogometasi$Player == "Ça?lar Söyüncü"] <- "Çağlar Söyüncü"
nogometasi$Player[nogometasi$Player == "Eldin Jakupovi?"] <- "Eldin Jakupović"
nogometasi$Player[nogometasi$Player == "Luka Milivojevi?"] <- "Luka Milivojević"
nogometasi$Player[nogometasi$Player == "Mat?j Vydra"] <- "Matěj Vydra"
nogometasi$Player[nogometasi$Player == "Mateo Kova?i?"] <- "Mateo Kovačić"
nogometasi$Player[nogometasi$Player == "Micha? Karbownik"] <- "Michał Karbownik"
nogometasi$Player[nogometasi$Player == "Nemanja Mati?"] <- "Nemanja Matić"
nogometasi$Player[nogometasi$Player == "Nikola Vla\u009ai?"] <- "Nikola Vlašić"
nogometasi$Player[nogometasi$Player == "Przemys?aw P?acheta"] <- "Przemysław Płacheta"
nogometasi$Player[nogometasi$Player == "Tomá\u009a Sou?ek"] <- "Tomáš Souček"

```

3. Pregled sezone

Ekipe koje su se natjecale u Premier Ligi u sezoni 2021/2022

```
as_tibble(klubovi)
```

```

## # A tibble: 20 x 1
##   value
##   <chr>
## 1 Arsenal
## 2 Aston Villa
## 3 Brentford
## 4 Brighton & Hove Albion
## 5 Burnley
## 6 Chelsea
## 7 Crystal Palace
## 8 Everton
## 9 Leeds United
## 10 Leicester City
## 11 Liverpool
## 12 Manchester City
## 13 Manchester United
## 14 Newcastle United
## 15 Norwich City
## 16 Southampton
## 17 Tottenham Hotspur
## 18 Watford
## 19 West Ham United
## 20 Wolverhampton Wanderers

```

```
klubovi
```

```

## [1] "Arsenal"           "Aston Villa"
## [3] "Brentford"         "Brighton & Hove Albion"
## [5] "Burnley"           "Chelsea"
## [7] "Crystal Palace"    "Everton"
## [9] "Leeds United"      "Leicester City"
## [11] "Liverpool"         "Manchester City"
## [13] "Manchester United" "Newcastle United"
## [15] "Norwich City"      "Southampton"

```

```
## [17] "Tottenham Hotspur"      "Watford"
## [19] "West Ham United"        "Wolverhampton Wanderers"
```

```
cat('Preko rownames mozda nastimati da ne pise "value" + odluciti jel bi ispisali preko tibble-a ili obicno')
## Preko rownames mozda nastimati da ne pise "value" + odluciti jel bi ispisali preko tibble-a ili obicno
```

```
najbolji_strijelci
```

Najbolji strijelci

```
##           Player           Team GlS GlS per 90 min
## 1      Mohamed Salah      Liverpool   23         0.75
## 2      Son Heung-min Tottenham Hotspur   23         0.69
## 3 Cristiano Ronaldo Manchester United   18         0.66
## 4       Harry Kane Tottenham Hotspur   17         0.47
## 5       Sadio Mané      Liverpool   16         0.51
```

```
najbolji_asistenti
```

Najbolji asistenti

```
##           Player           Team Ast Ast per 90 min
## 1      Mohamed Salah      Liverpool   13         0.42
## 2 Trent Alexander-Arnold      Liverpool   12         0.38
## 3       Mason Mount      Chelsea   10         0.38
## 4       Harvey Barnes Leicester City   10         0.43
## 5      Andrew Robertson      Liverpool   10         0.35
## 6       Jarrod Bowen West Ham United   10         0.30
```

Pozicije igrača izbrisati Po ovome vidimo da nemamo klasicne pozicije nego imamo vezni, napadaci, napadaci-vezni, vezni-obrambeni i slicne kombinacije

```
table(factor(nogometasi$Pos)) %>% sort()
```

```
##
## FW,DF DF,FW DF,MF MF,DF MF,FW FW,MF GK FW MF DF
##      2      4      9      11      39      62      74      99      158      233
```

```
samoVezni <- nogometasi %>% filter(Pos == "MF")
samoNapadaci <- nogometasi %>% filter(Pos == "FW")
samoObrambeni <- nogometasi %>% filter(Pos == "DF")
golmani <- nogometasi %>% filter(Pos == "GK")
vezniNapadaci <- nogometasi %>% filter(Pos == "MF,FW" | Pos == "FW,MF")
vezniObrambeni <- nogometasi %>% filter(Pos == "MF,DF" | Pos == "DF,MF")
obrambeniNapadaci <- nogometasi %>% filter(Pos == "DF,FW" | Pos == "FW,DF")
nrow(samoVezni)
```

```
## [1] 158
```

```
nrow(samoObrambeni)
```

```
## [1] 233
```

```
nrow(samoNapadaci)
```

```
## [1] 99
```

```
nrow(vezniNapadaci)
```

```
## [1] 101
```

```
nrow(vezniObrambeni)
```

```
## [1] 20
```

```
nrow(obrambeniNapadaci)
```

```
## [1] 6
```

```
nrow(golmani)
```

```
## [1] 74
```

```
mladi <- nogometasi %>% filter(Age < 25)
nrow(mladi)
```

Godine igrača

```
## [1] 355
```

```
stari <- nogometasi %>% filter(Age >= 25)
nrow(stari)
```

```
## [1] 332
```

4. XY-test

Prvi od testova koje ćemo koristiti je xy-test. Njega smo odlučili koristiti jer **insert__razlog**. Tim testom želimo odgovoriti na pitanja kao što su:

- Postoji li razlika u broju odigranih minuta mladih igrača (do 25 godina) među premierligaškim ekipama?
- **sl pitanje**

Test ćemo izvesti za **1. i 2. pitanje (dogovorimo)** te ćemo pojasniti potrebne korake. ...

Zaključci: ...

5. YZ-test

Nakon prvog testa, sljedeći po redu je yz-test. Pitanja na koja nam on može dati odgovor su recimo:

- Dobivaju li u prosjeku višestrukih žutih kartona napadači ili igrači veznog reda?
- **sl pitanje**

On nam je ovdje koristan jer **dopuniti...** ...

Zaključci: ...

6. ZX-test

Što je zapravo uspješnost igrača? To je pitanje kojim smo se prvotno morali baviti i secirati što čini dobrog igrača ovisno o pozicijama. Došli smo do sljedećih zaključaka:

- **nesto nesto**
- **nesto nesto**
- **nesto nesto**

- nesto nesto

blablabla ...

Zaključci: ...

7. XZ-test

Svi koji prate nogomet malo detaljnije znaju čiji igrači se cijene. Brazilci su najbolji dribleri, Španjolci najbolji u tiki-taki, Hrvati najbolji u penalima, ali u Engleskoj su najbolji Englezi. Javnost to zove “*English tax*” i time se cilja na činjenicu kako engleski klubovi skuplje plaćaju i prodaju domaće igrače u odnosu na strane. Je li to opravdano, pokazat će nam naš xz-test. Koristit ćemo ga jer **razlog** i napokon ćemo saznati doprinose li oni sveukupnom uspjehu tima ili je to još jedna preuveličana engleska nogometna bajka. ...

Zaključci: ...

8. Rezultati

...

...