

Úvod

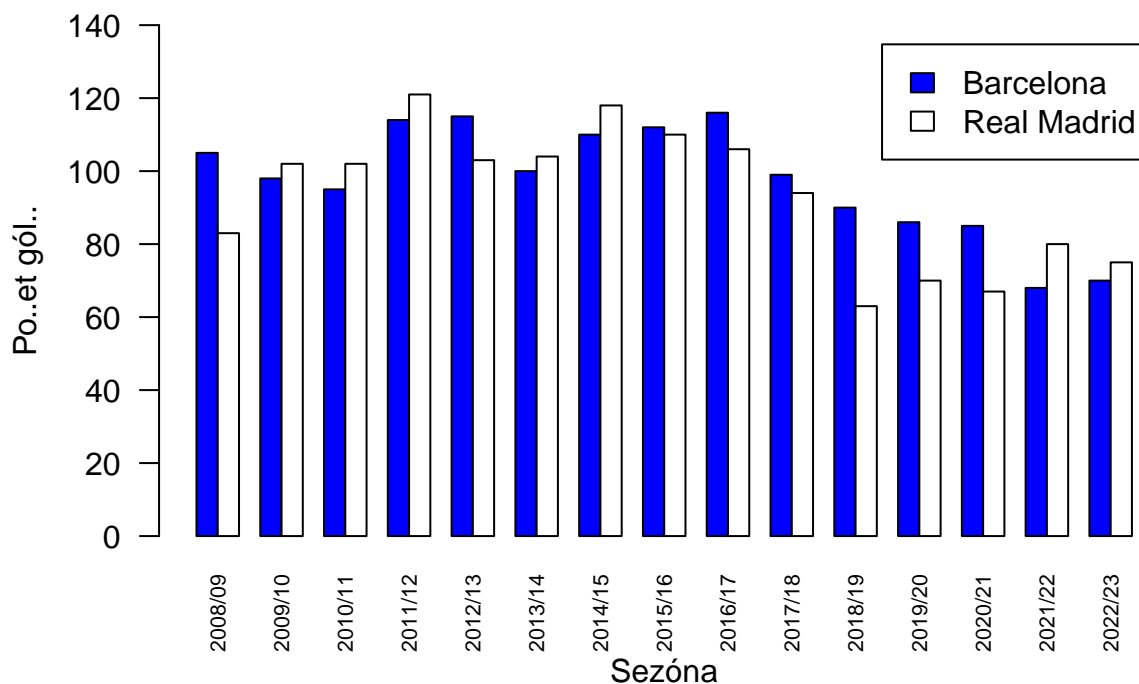
V mé semestrální práci jsem se rozhodl využít párový t-test pro porovnání vstřelených gólů za celou sezónu mezi dvěma největšími španělskými fotbalovými kluby, FC Barcelona a Real Madrid. Data jsou od sezóny 2008/09 až po loňskou sezónu 2022/23. (Aktuální sezóna 2023/24 ještě není ukončena je odehráno 29 z 38 zápasů a tudíž by data byla zkreslená). Do dat jsou zahrnuty pouze góly ze zápasů La ligy a tedy nejsou zahrnuty góly z jiných soutěží (Copa del Rey, Copa de España, Liga Mistrů, Superpohár atd.).

Data jsem čerpal ze stránky [footystats](https://footystats.com/), kde jsou přehledně zobrazeny.

Sezóna	Barcelona	Real Madrid
2008/09	105	83
2009/10	98	102
2010/11	95	102
2011/12	114	121
2012/13	115	103
2013/14	100	104
2014/15	110	118
2015/16	112	110
2016/17	116	106
2017/18	99	94
2018/19	90	63
2019/20	86	70
2020/21	85	67
2021/22	68	80
2022/23	70	75

Vizualizace dat

Počet vstřelených gólů za sezónu



Definice základní pojmů

Statistický znak x:

- Počet gólů vstřelených za sezónu

Statistická jednotka:

- FC Barcelona
- Real Madrid

Základní soubor:

- všechny sezóny španělské ligy

Výběrový soubor:

- 15 sezón španělské ligy (2008/09 - 2022/23)

Výpočet základních číselných charakteristik

Barcelona

Průměr

```
mean(barcelona)
```

Medián

```
median(barcelona)
```

Rozptyl

```
var(barcelona)
```

Směrodatná odchylka

```
sd(barcelona)
```

Real Madrid

Průměr

```
mean(real_madrid)
```

Medián

```
median(real_madrid)
```

Rozptyl

```
var(real_madrid)
```

Směrodatná odchylka

```
sd(real_madrid)
```

Shrnutí

	Barcelona	Real Madrid
Průměr	97,53333	93,2
Medián	99	102
Rozptyl	237,8381	353,4571
Směrodatná odchylka	15,422	18,80046

Ověření podmínky normality (sahpira-wilkův test)

H_0 : Sledovaný znak má normální rozdělení

H_1 : Sledovaný znak nemá normální rozdělení

$\alpha = 0,05$ hladina významnosti

Barcelona

```
shapiro.test(barcelona)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: barcelona  
## W = 0.92152, p-value = 0.2032
```

P-hodnota je 0,2032 a tedy $P_{hodnota} > \alpha = 0,2032 > 0,05$ a ponecháváme nulovou hypotézu. Což znamená, že znak má normální rozdělení.

Real Madrid

```
shapiro.test(real_madrid)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: real_madrid  
## W = 0.93144, p-value = 0.2867
```

P-hodnota je 0,2867 a tedy $P_{hodnota} > \alpha = 0,2867 > 0,05$ a ponecháváme nulovou hypotézu. Což znamená, že znak má normální rozdělení.

Párový t-test

$H_0 : \mu_A = \mu_B$ (celkový průměrný počet gólů za sezonu je stejný pro oba týmy)

$H_1 : \mu_A \neq \mu_B$ (celkový průměrný počet gólů za sezonu není stejný pro oba týmy)

$\alpha = 0,05$ hladina významnosti

kde:

μ_A je průměrný počet gólů za sezonu pro Barcelonu

μ_B je průměrný počet gólů za sezonu pro Real Madrid

```
t.test(barcelona,real_madrid,paired=TRUE)
```

```
##  
## Paired t-test  
##  
## data: barcelona and real_madrid  
## t = 1.3551, df = 14, p-value = 0.1968  
## alternative hypothesis: true mean difference is not equal to 0  
## 95 percent confidence interval:  
## -2.525088 11.191755  
## sample estimates:  
## mean difference  
## 4.333333
```

P-hodnota je 0,1968 a tedy $P_{hodnota} > \alpha = 0,1968 > 0,05$ a ponecháváme nulovou hypotézu. Což znamená, že průměrný počet gólů za sezonu je stejný pro oba týmy.

Závěr

Na základě provedeného t-testu, p-hodnota je 0,1968, což je větší než standardní hladina významnosti $\alpha = 0,05$. To znamená, že neexistuje dostatečný statistický důkaz pro zamítnutí nulové hypotézy. Nulová hypotéza, která předpokládá, že průměrný počet gólů vstřelených Barcelonou a Realem Madridem za sezonu je stejný. Jinými slovy, na základě dostupných dat a provedeného t-testu, nemůžeme tvrdit, že by existoval statisticky významný rozdíl v průměrném počtu gólů vstřelených těmito dvěma týmy za sezonu.