Football betting data

Roudranil Das

2022-09-05

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data <- read.csv("E0.csv", header = T)
## FTHG: full time home goals
FTHG <- data[,"FTHG"]
freq_table <- table(FTHG)
freq_table
```

```
## FTHG
##   0   1   2   3   4   5   6   7
##  93 121  89  50  15   8   2   2
```

Using this data we would like to find out the probability that in 22-23, there will be a match where the home team will score 8 or more goals?

Based on this data, we would say that the prob is 0, but that is not the case. Lets model using a distribution. Candidates:
1. Poisson
2. Geometric
3. Negative binomial

Let $X : FTHG$. Then say assume $X \sim Poisson(\lambda)$. Then we want to find out

$$P[X \geq 8] = 1 - P[X \leq 7]$$

Using method of moments
We know that $E[X] = \lambda$. Sample mean is a good guess for the population mean. Thus

```
lambda_hat = mean(FTHG)
lambda_hat
```

```
## [1] 1.513158
```