Predicting with models and distributions

Narek Sahakyan

September 25, 2019

Overall 50pts. Submit both markdown and pdf files You were right Mr. Madoyan, I started on wednesday.

I am Narek Sahakyan, and I am assigned to analyze the Dutch league(Eredivisie). Before starting the analysis I know that, Eredivisie is considered one of the most attacking leageues in the world, so let's see if this statement is expressed in their statistics.

```r
library(SportsAnalytics270)
data(f_data_sm)
```

1. Calculate average number of goals for home and away teams per SEASON for that league, use dplyr. (4p)

```r
eredivisie <- f_data_sm %>%
  filter(LEAGUE == "Eredivisie")

home_goals <- eredivisie %>%
  group_by(SEASON, HOMETEAM) %>%
  summarise(HOME_MEAN = mean(FTHG))

away_goals <- eredivisie %>%
  group_by(SEASON, HOMETEAM) %>%
  summarise(AWAY_MEAN = mean(FTAG))

#binding the results into a data frame for later usage
mean_goals <- home_goals
mean_goals <- rename(mean_goals, TEAM = HOMETEAM)
mean_goals$AWAY_MEAN <- away_goals$AWAY_MEAN
```

2. Construct a plot using ggplot to show how this number is changing over time. Comment Note: you need to have SEASON on x-axis. Show average Home goals and Away goals on the same plot. Be sure that your plot has appropriate axis names and title.(4p)

3. Interpret the plot (4p)

4. Think of your own approach on how will you measure home team advantage given the data you have. Calculate Home team advantage for your league and for all other leagues over time (4p).

5. Plot the results of the exercise 4 here using ggplot2 and interpret it (4p).

Predictions

1. With your chosen league, predict the probabilities of the first week of the season. Use distribution approach. (4p)

Probabilities

2. Now use Poisson regression for the same task as above. Create the model and predict the probabilities. Interpret the home team coefficient. (6p)

Data preperation

Model

Probalities

3. Compare the probabilities to each other and to the actual results. How good do you think you were? Can you think of a metric that can help us to understand how good are our predictions over the week (2p)

Bradley terry model

You need to construct Bradley-Terry model for NBA regular season games. Dataset is nba2009_2018 from SportsAnalytics270 package

Remove the game Boston Celtics vs Indiana Pacers from the dataframe. This is a game that was cancelled due to Boston Marathon bombing

1. Do the Data Prep here (4p)

2. Model here (4p)

3. Plot the abilities: Which team is the beast, which the worst ? (4p)

4. Make prediciton for the first 3 games of the new season (Schedule can be found here: https://tinyurl.com/y4hpbl8f) (6p)