

# Final exam

*Michael Lopez, Skidmore College*

Note: Please submit Section 1 of the exam using RMarkdown. Section 2 can use either Microsoft Word or RMarkdown. The exam is due Thursday at noon; seniors, please make an effort to get it in on Wednesday.

Unless there are extenuating circumstances, the exam must be printed, stapled, and submitted to my mailbox. 5 points are awarded based on presentation, grammar, and syntax.

## Section 1: Team rankings (45 points. Q5 worth 15)

The goal of this section is to implement the Bradley-Terry model to estimate team strength during the current season (2015-16) English Premier League.

First, here's some code to extract the data, identify whether or not the home team won, and run the Bradley Terry model (BTM).

```
epl.df <- read.csv('http://www.football-data.co.uk/mmz4281/1516/E0.csv')
epl.df <- epl.df %>%
  select(Date, HomeTeam, AwayTeam, FTHG, FTAG) %>%
  mutate(Outcome = ifelse(FTHG > FTAG, 1, ifelse(FTHG < FTAG, 0, 0.5)))
head(epl.df)
homeBT <- BTM(Outcome,
              data.frame(team = HomeTeam, home.adv = 1),
              data.frame(team = AwayTeam, home.adv = 0),
              ~ team + home.adv,
              id = "team", data = epl.df)
```

1. According to the BTM, identify the top 5 teams in the current EPL season.
2. Does there appear to be a statistically significant difference in team ability coefficients between Leicester and Manchester City?
3. Estimate the probability that Leicester beats Manchester City (i) at home (ii) at a neutral site and (iii) on the road.
4. Interpret the estimated home advantage coefficient (on the odds scale) from the BTM output.
5. Has the home advantage in the EPL increased, decreased, or stayed the same? Data from past seasons can be obtained by changing part of the url given above. Specifically, 1516 refers to the year, and so the web address `http://www.football-data.co.uk/mmz4281/0506/E0.csv` would access the 0506 season. Data goes back through the 1995-96 season. Review past seasons to identify if the home advantage in the EPL has increased, decreased, or if evidence is inconclusive. You do not need to review all seasons, but use at least two other seasons (feel free to look at more, too) in your response. Your response should look at both (i) the fraction of games won by the home team and (ii) the coefficient on the home advantage term in a BTM model.

## Section 2 (50 points)

One of the goals of our course was to be able to communicate more complex statistical developments in a manner that those unfamiliar with statistics can understand. In this section, you will write a roughly two-page summary designed for a team official that is both interpretable and actionable. Your job is to summarize recent work that's been done, explain why and how any traditional statistics may be flawed, as well as to comment on unsolved issues. Ideally, the team official (coach or general manager) could use your comments to understand past team success, develop strategies for future games, and/or acquire new players.

Pick exactly one of the following three sports - **baseball, basketball, or hockey**. You may refer to our readings, class notes, or outside material in your write-up.