WSABI Summer Research Lab **Summer 2025**

Lab 14: Fully Bayesian Models

Instructor: Jonathan Pipping *Authors: RB, JP*

14.1 Home Field Advantage in the NFL

In 2017, Mike Lopez and Gregory Matthews published a paper about the randomness in North American sports [LMB]. As part of this paper, they created the plot in Figure 14.2, which shows the differences in home field advantage across every franchise of the four major sports leagues in the United States.

In this lab, you will replicate this plot for NFL teams, using game-level data from the 2018-2023 seasons.

14.1.1 Data

As in Lecture 14, we will use data from NFL games from 2018-2023, which is displayed in Figure 14.1.

game_id	home_team	away_team	season_type	week	total_home_score	total_away_score	season	pts_H_minus_A
2018_01_ATL_PHI	PHI	ATL	REG	1	18	12	2018	6
2018_01_BUF_BAL	BAL	BUF	REG	1	47	3	2018	44
2018_01_CHI_GB	GB	CHI	REG	1	24	23	2018	1
2018_01_CIN_IND	IND	CIN	REG	1	23	34	2018	-11
2018_01_DAL_CAR	CAR	DAL	REG	1	16	8	2018	8
2018_01_HOU_NE	NE	HOU	REG	1	27	20	2018	7
2018_01_JAX_NYG	NYG	JAX	REG	1	15	20	2018	-5
2018_01_KC_LAC	LAC	KC	REG	1	28	38	2018	-10
2018_01_LA_OAK	LV	LA	REG	1	13	33	2018	-20
2018_01_NYJ_DET	DET	NYJ	REG	1	17	48	2018	-31
2018_01_PIT_CLE	CLE	PIT	REG	1	21	21	2018	0
2018_01_SEA_DEN	DEN	SEA	REG	1	27	24	2018	3

Figure 14.1: First few rows of the NFL game-level data.

To make your modeling easier, add the following columns to the data frame:

- $S(i)$: the season the i^{th} game took place, indexed at 1 for the 2018-2019 season
- $H(i), A(i)$, indexes for the home and away teams (respectively) in the i^{th} game
- y_i the home-minus-away score differential ($y_{H_i} - y_{A_i}$) in the i^{th} game

14.1.2 Your Task

1. When modeling home field advantage, it is essential to include team strength. Modify the model from Lecture 14 to include team-specific home field advantage parameters; write this model out and fit it using `Stan` in `R`. Instructions for how to do this are provided below, but try to do it on your own first.
2. Plot the posterior means of the home field advantage parameters for each team, along with their 95% credible intervals. Which teams have the largest home field advantage? Which have the smallest?

3. Plot the posterior means of the team strength parameters for each team, along with their 95% credible intervals. Which teams are the strongest? Which are the weakest?

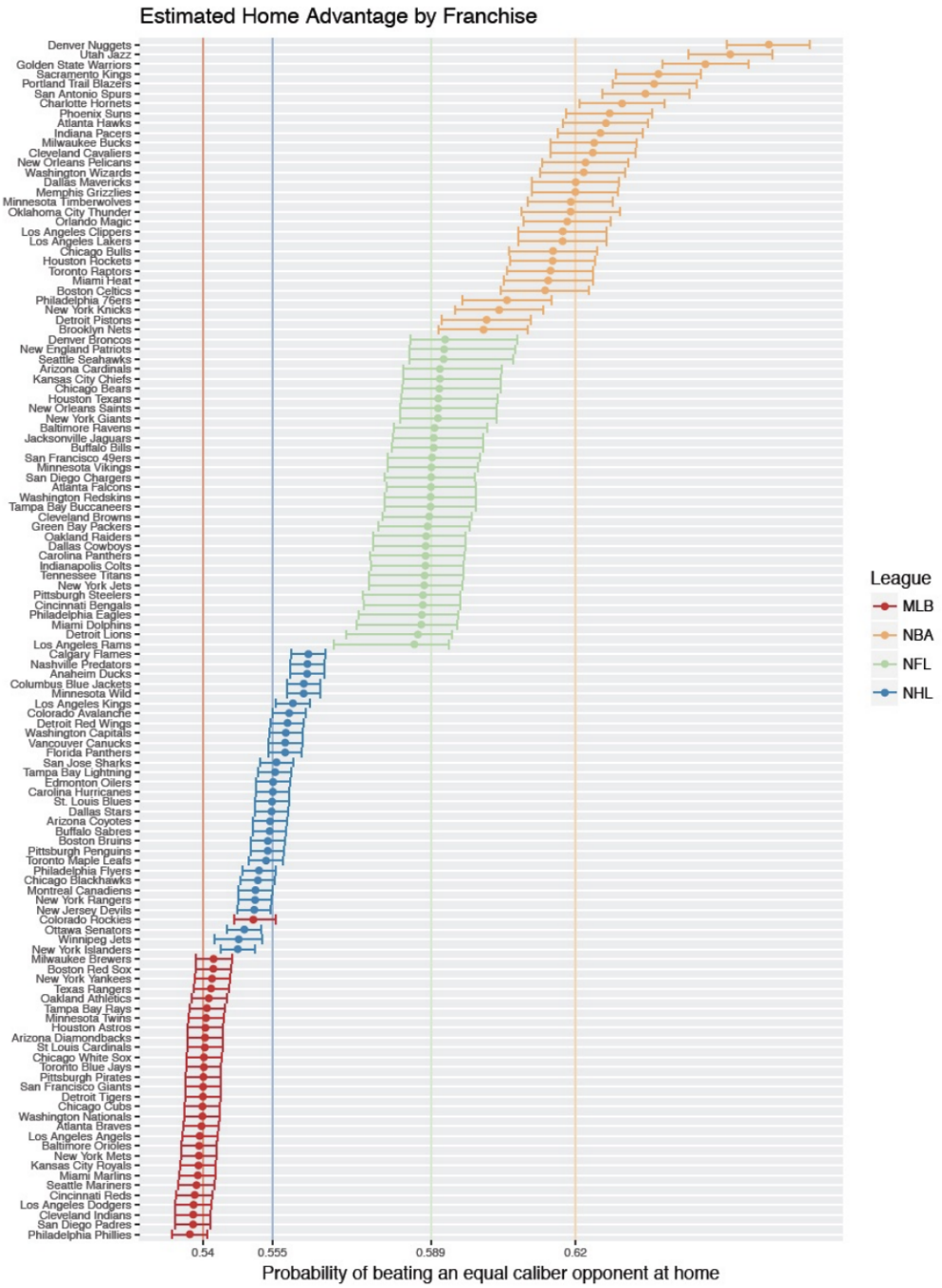

Figure 14.2: Posterior means and 95% credible intervals for the home field advantage parameters for each team in the four major US sports leagues.

14.1.3 Setting Up the Model

We define team-specific and overall home field advantage parameters as follows:

- α_j: the home field advantage parameter for team j
- α: the overall home field advantage parameter

We have to define priors for these parameters, and we do so as follows:

$$\begin{aligned} \alpha_j &\sim \mathcal{N}(\alpha, \sigma^2_{\text{HFA}}) \\ \alpha &\sim \mathcal{N}(0, \tau^2_{\text{HFA}}) \\ \sigma^2_{\text{HFA}} &\sim \mathcal{N}_+(0, 5^2) \\ \tau^2_{\text{HFA}} &\sim \mathcal{N}_+(0, 5^2) \end{aligned}$$

The code to fit this model using `Stan` in `R` is demonstrated in Lecture 14.

References

[LMB] Lopez, M. J., Matthews, G. J., and Baumer, B. S., *How often does the best team win? A unified approach to understanding randomness in North American sport*, Research Article, 2017.