

# STA 610: HW 4

*Graham Tierney*

*November 17, 2019*

## Introduction

In the National Football League (NFL) and many other sports leagues, each team does not play every other team. This leaves many fans wondering how their team would match up against others who were never played, particularly in playoffs. As such, I will build a tool using hierarchical models to predict the scores (and outcomes) of football games using data from the 2018 NFL season. I include all games, regular season and playoffs, and focus the final model predictions on who among the final four teams (Kansas City Chiefs, New England Patriots, New Orleans Saints, and LA Rams) were most likely to win the Super Bowl.

## Model

Each game has a two-dimensional outcome.  $Y_{hg}$ , the score of the home team  $h$  in game  $g$  and  $Y_{ag}$ , the score of the away team  $a$  in game  $g$ . Modeling the joint distribution of  $(Y_{hg}, Y_{ag})$  using the tools we've studied in class would be quite difficult. A multivariate normal wouldn't be quite appropriate because observed scores are often close to zero but must be non-negative. The conditional distribution  $Y_{hg}|Y_{ag}$  is also likely not normal for the same reasons, and is potentially very complex.

To predict the outcome for a game, however, all we need to estimate is the score differential,  $Y_{hg} - Y_{ag}$ . This quantity can take on positive and negative values and has positive support on nearly all integers between -100 and 100. I will assume that the home-team score differential is distributed as:

$$Y_{hg} - Y_{ag} \sim N(\beta_0 + \mu_h - \mu_a, \sigma^2)$$

$\mu_h$  and  $\mu_a$  are latent variables measuring the skill of the home and away teams.  $\beta_0$  captures the concept of "home-field advantage." If the home and away teams are equally skilled, the home team is favored by  $\beta_0$  points. Note that the Super Bowl location is picked regardless of the teams playing in the game, so the "home" and "away" designations are arbitrary. Consequently, I only include  $\beta_0$  in regular season and non-Super Bowl playoff games.

The main alternative to the normal assumption on the point differential would be using count models. In sports where each score is worth the same amount of points (soccer or baseball, for example), count models could be used to identify the marginal distributions of  $Y_{tg}$  (the score for team  $t$  in game  $g$ ). However, in football points are accrued in increments of 2, 3, 6, 7, or 8. As such, a count model would need to have at least a five-dimensional outcome for each team. This would further increase the model complexity and make it more difficult to specify hierarchical models similar to ones covered in class. The normal assumption here can be thought of as an approximation to the true model.

The hierarchical component of the model comes in modeling  $\mu_t$ . I assume  $\mu_t \sim N(\mu_0, \tau^2)$ . The hierarchical assumption is quite useful because NFL seasons consist of only 16 games per team. This is not a lot of data to estimate a team-specific parameter. However, because the sampling model for  $Y_{hg} - Y_{ag}$  is a function of the pairwise-differences, the scale of each  $\mu_t$  is not identified. Adding a constant to each  $\mu_t$  produces an equivalent data likelihood value. As such, I set  $\mu_0 = 0$  and assume that  $\sum_t \mu_t = 0$ . This is different from the traditional approaches in class where we often are interested in the grand mean  $\mu_0$ . While one might be interested in estimating something like the expected score across teams, this model cannot provide such estimates.

## Estimation

I estimate the parameters using JAGS and place diffuse inverse-Gamma priors on  $\sigma^2$  and  $\tau^2$  and a diffuse normal prior on  $\beta_0$ . I also compare the results from this Bayesian framework to the OLS values for  $\mu_t$  estimated without any shrinkage. The model above (excluding the heirarchical assumption) can be expressed as  $\mathbf{Y} \sim N(\beta_0 \mathbf{1} + \mathbf{X}\boldsymbol{\mu}, \sigma^2)$ , where  $\mathbf{Y}$  is the vector of home team score differentials,  $\mathbf{1}$  is a vector of 1s,  $\boldsymbol{\mu}$  is the vector of team parameters  $\mu_t$ , and  $\mathbf{X}$  is the design matrix.  $X_{it} = 1$  if team  $t$  is the home team in game  $i$ ,  $-1$  if  $t$  is the away team, and 0 otherwise. The OLS and Bayesian estimates of  $\mu_t$  are shown below, sorted by the OLS estimate.<sup>1</sup>

Team	Bayes	OLS	Team	Bayes	OLS
NO	6.26	9.35	ATL	-0.28	-0.28
KC	6.12	8.89	CLE	-0.90	-0.59
LA	5.66	7.94	DEN	-0.69	-0.73
NE	5.47	7.30	NYG	-1.63	-2.16
BAL	4.05	6.33	GB	-1.43	-2.63
CHI	4.68	6.01	TB	-2.35	-2.74
LAC	3.94	5.74	DET	-1.70	-2.90
PIT	3.41	5.46	CIN	-2.99	-3.61
SEA	3.03	4.18	JAX	-2.61	-3.77
IND	2.99	3.84	WAS	-3.36	-4.91
HOU	2.55	3.22	SF	-3.91	-5.71
PHI	1.38	2.30	BUF	-4.36	-6.59
DAL	0.75	1.32	NYJ	-4.85	-7.50
CAR	0.11	0.76	MIA	-5.40	-8.48
MIN	0.78	0.69	OAK	-6.89	-9.47
TEN	0.37	0.43	ARI	-8.18	-11.67

The top four teams are the same under OLS and the heirarchical estimation, and appear to be a cut above the rest. They were the final four teams in the 2018 season; they all played in the conference championship games. The New England Patriots won the Super Bowl, but, according to the estimation, were the worst of the final four teams. The Bayesian model provides a nice way to measure how likely their win was and how the other teams would have fared in the Super Bowl. Each draw of the parameters in the Gibbs sampler can be plugged into the CDF of  $Y_{hg} - Y_{at}$  to find  $P(Y_{hg} - Y_{at} > 0)$ , the probability that the home team wins. I ignore  $\beta_0$  in this case because the Super Bowl is not a home game for either team.

Team 1	Team 2	$\mu_1$	$\mu_2$	P(Team 1 Wins)	95% CI
LA	KC	4.95	5.42	0.49	(0.28,0.70)
NE	KC	4.62	5.42	0.48	(0.27,0.68)
NO	KC	5.46	5.42	0.50	(0.29,0.70)
NE	LA	4.62	4.95	0.49	(0.28,0.69)
NO	LA	5.46	4.95	0.52	(0.30,0.72)
NO	NE	5.46	4.62	0.53	(0.31,0.73)
KC	ARI	5.42	-8.86	0.86	(0.69,0.96)
LA	ARI	4.95	-8.86	0.85	(0.69,0.95)
NE	ARI	4.62	-8.86	0.85	(0.67,0.95)
NO	ARI	5.46	-8.86	0.86	(0.69,0.96)

The  $\mu_1$ ,  $\mu_2$ , and P(Team 1 Wins) columns report the posterior means for the indicated quantities. The 95% CI is for the probability. While the model ranks the Saints as the best team, they lost to the LA Rams in the conference championships. The probability that the Saints would have defeated the Patriots in the Super Bowl is about 53% (95% CI from 31% to 73%). The Rams lost the Super Bowl, but the outcome if the Saints

<sup>1</sup>The identification issue is resolved by dropping one of the teams from  $\mathbf{X}$  (setting  $\mu_t = 0$  for one  $t$ ). To compare to the Bayesian estimates where  $\sum_t \mu_t = 0$ , I impute 0 for the missing OLS estimate and then subtract the mean. This ensures the OLS estimates follow the same restriction as the Bayesian ones.

had made it to the final game would have been essentially a coinflip. All of the top teams were quite close in skill.

I also included the potential matchups against the Arizona Cardinals to demonstrate that the model does make some extremely confident predictions. The top four teams all had about an 85% probability of beating the Cardinals, ignoring home field advantage even though the Cardinals had no chance to make the Super Bowl.

## Conclusion

Hierarchical models provide a useful way to borrow strength across groups with small sample sizes. While the 2018 NFL season consisted of 267 games, each team played only 16 to 19 games (depending on post-season success). In each game, two teams were present, so the outcome is a function of the interaction between both team's skill. The simple OLS model then estimates 32 mean parameters ( $\beta_0$  and  $\mu_t$  for 31 teams) with only about 8 observations per parameter. A hierarchical model is useful then to share information across teams.

The other interesting feature of this application is that  $\mu_0$ , which we usually want to estimate, is actually nearly impossible to estimate in this setup. Setting  $\mu_0$  to 0 is a very reasonable assumption in this case.

The model could also be extended to add further levels of hierarchy. One could add multiple seasons of data to the model by placing another level of hierarchy on  $\mu_t$ . Let  $s$  index the seasons in the data. A plausible extension to the sampling model is:  $Y_{hsg} - Y_{asg} \sim N(\beta_0 + \mu_{hs} - \mu_{as}, \sigma^2)$ ,  $\mu_{ts} \sim N(\mu_t, \phi^2)$ , and  $\mu_t \sim N(0, \tau^2)$ . Here the model is the same within a season, team  $t$ 's skill in season  $s$  is  $\mu_{ts}$  and for all  $s$ ,  $\mu_{ts}$  are iid draws from a normal distribution with a team-specific mean  $\mu_t$ . I did not include this extension because the NFL has such a significant churn in which teams are good year to year; the bottom teams are rewarded with high draft picks, who can quickly turn a team around. Thus, I expect (and found) that  $\phi^2$  was quite large.

A further random effect could be included with  $\beta_0$  to identify if certain teams have a larger home field advantage. The Chargers and Rams recently moved to LA from San Diego and St. Louis, so their home field advantage could be smaller. I did not want to add another 32 parameters to the model to estimate and I was focused on the Super Bowl match ups, so I did not include that extension in this application. It would be relatively easy to implement, however, and could be interesting.