

**THE SUPER BOWL:
PREDICTING THE TEAMS**

Skyler Johnson
Sarah Witt
Jamie Adams

BAYESIAN STATISTICS
PROFESSOR COWLES

INTRODUCTION

For our report, we wanted to establish a model that would allow us to predict a teams' probability of making it to the Super Bowl based on their regular season statistics. We started by analyzing NFL data for a ten year period, beginning in 1997 and ending in 2006. The analysis included gathering the following data for each team: passing yards per game, rushing yards per game, passing yards allowed per game, rushing yards allowed per game and turnover margin. Turnover margin is defined as total turnovers forced (interceptions, fumbles, etc.) minus turnovers allowed; a positive turnover margin is desirable. These five statistics were gathered for every NFL team. We split the NFL teams up according to their conference, AFC or NFC. This was performed in case the two conferences had different significant values. For example, the NFC could be rush heavy with the best teams having the best rushing games, while the AFC could be a defensive conference with the top teams having the best defenses. We were also interested in whether there was a significant trend in the type of teams that were making the Super Bowl. For example, teams in the 90's may have been more run-oriented, while teams in more recent years could emphasize the passing game.

METHODS

We gathered our data from NFL.com. We then standardized each statistic by conference and year. For each year, we analyzed the data twice using WinBUGS. The first analysis involved using a non-informative prior and the second was with an informative prior. To get the informative priors we fit a logistic regression model into SAS. The AFC SAS output was used as the informative priors for the NFC WinBUGS analysis and the NFC SAS output was the

informative priors for the AFC analysis. Using an informative prior should give us significantly better data.

In addition to the five previously mentioned variables, we included a variable to account for statistical dependence across years for a particular team. Each team was given a unique number that remained constant regardless of the year (i.e. Arizona was assigned number one for all ten years).

We started every analysis with a full logistic regression model and removed the insignificant effects one at a time to ensure the best results. Using WinBUGS, we checked for convergence and autocorrelation for each of the parameters. These plots are shown in the Analysis, as well as, in the Appendix. After deciding on a model, Super Bowl probabilities for 1996 and 2007 were predicted. Prediction for 1996 gave the advantage of allowing us to check the accuracy of the model.

For the second part of our analysis, we fit a simple linear regression model in WinBUGS for each statistic across year. Output showing a slope coefficient significantly different from zero would indicate a positive change in the statistic across years. For example, an analysis of AFC passing offense resulting in a positive slope would signify a change in strategy towards a more pass oriented offense over the past ten years.

ANALYSIS

We began in WinBUGS with an analysis of the AFC using non-informative priors. Our initial model (Appendix A) included all six variables. Our output indicated Passing Defense and the Team variable were not significant. After removing these variables the model was run again. (Appendix A) All remaining variables were significant.