Aleena Eskeets, Sam Lee
Dr. Berrett
STAT 251
16 November 2023

Predictive Model for March Madness Wins

In past years, the winning and losing scores of teams that have played in March Madness appear to have a gamma distribution. Given the γ and ϕ of the winning and losing scores, as well as various other covariates (turnovers, offensive rebounds, defensive rebounds, field goal percentage, free throw percentage, and three-point percentage for both $Team_i$ and $Team_j$), can we predict whether $Team_i$ or $Team_j$ will win in a game?

Essentially, we are trying to find the posterior distribution of $Score_{ij}|data$, where $Score_{ij}$ is the ith team's score against $Team_j$.

Prior: $Score_{ij} \sim Gamma(\gamma, \phi)$

Predictive Posterior: $(Score_{ij} > Score_{ji}|data) \sim Bernoulli(p_{ij})$

Additionally, each covariate for $Team_i$ and $Team_j$ will have its own posterior distribution estimates: turnovers, offensive rebounds, and defensive rebounds will all have gamma conjugate prior distributions, while field goal percentage, free throw percentage, and three-point percentage will all have beta conjugate prior distributions.

We want to estimate p_{ij}, the probability $Team_i$ will win against $Team_j$—can we predict the outcome of any given NCAA Division I matchup controlling for these effects?

Our sources of data can be found on:

1. March Madness Learning Mania 2023: https://www.kaggle.com/competitions/march-machine-learning-mania-2023
2. 2024 Customizable T-Rank and Tempo-Free Stats: https://www.barttorvik.com/#