

ST 758 Project 3

Jessica Miller

Dec. 1, 2016

In this assignment, I used a 64-seed bracket to predict the winner of a tournament. Mainly, I followed the formula defined in the project guidelines: $\text{logit}[P(\text{seed } i \text{ beats seed } j)] = \beta(\sqrt{j} - \sqrt{i})$.

I chose to focus my efforts on the parameter of the above function, β . Adjusting this value changes the probability that a certain seed will be chosen over, or "defeat", its opponent. After combining research and prior knowledge of the NCAA March Madness tournament, I divided my 1000 prediction opportunities into four cases. In each case, I investigate the percentage of the time the number one seed is chosen to win the whole tournament. As the top seed, one would think it wins the majority of the time, but as is evident in every March Madness tournament, the top seed is upset quite often. My efforts in this project focused on these upsets.

The first case of predictions I chose uses the simple parameter outlined in the project description: 0.6. I used 0.6 as a starting value for the parameter in order to gain a sense of how the tournament would play out initially. The first 250 columns in the "jmmill13ST758.Rdata" file's brackets matrix follow this case. According to my resulting matrix, the average winning seed of these 250 predictions is 3.48. The lowest seed predicted to win is 18, and the number one seed wins $\frac{67}{250} = 27\%$ of the time.

The second case of predictions I chose assumes the probability the higher seed defeats the lower seed is smaller than initially thought. Ideally, the top seeds would lose more often. In the context of this setup, these probabilities attempt to model a smoother tournament; that is, the top seeds carry less weight in their success. I chose $\beta = 0.5$ to predict the next 250 tournaments, or lines 251-500 in the matrix. Under this restraint, the average winner is seed 4.63. The lowest seed predicted to win is 25 and the number one seed wins $\frac{58}{250} = 23\%$ of the time.

The third case of predictions becomes slightly more complex. During the March Madness tournament, typically there are fewer upsets in the first few rounds. However, as the tournament progresses, the lesser teams are weeded out and the winner of each game is not as definite. In this model, the upsets occur more often as the rounds increase. I set the parameter to its initial 0.6 and decreased it by 0.025 every round, lessening the probability the higher seed defeats the lower seed. From matrix columns 501-750, the average winning seed is 4.43 and the lowest seed chosen to win is 24. The percentage of number-one-seed wins is $\frac{60}{250} = 24\%$.

The final case, and last 250 columns, of predictions that I chose represents what I like to call the "underdog effect". It is common knowledge that the March Madness tournament is filled with upsets galore; in fact, FiveThirtyEight uses the statistics and probabilities behind upsets to help model the tournament (see <http://fivethirtyeight.com/features/how-to-tell-if-a-march-madness-underdog-is-going-to-win/>). I tried to calculate this in the tournament with a seeding threshold: if the two teams playing are less than or equal to 20 seeds away from each other, the probability the higher seed defeated the lower seed is lessened. I chose to decrease the parameter by 0.15 to make the probability more even. However, in my code and brackets matrix that I turned in, I accidentally increased the

parameter by 0.15, which gave me an average winning seed of 2.94. This is the opposite of what I was intending. For the sake of this write-up, I re-ran the calculations using the parameter I intended and obtained more logical results. Here, the average winning seed is 4.55 and the number one seed wins $\frac{67}{250} = 27\%$ of the time. Although the top seed wins the same number of times as in the first case, the average winner has a lower seed, implying there are more lower seeds who win than in case one.

To summarize my results, I created Table 1 below:

Table 1: Summary results for each case of tournament predictions.

Case	Average Winner	Percentage 1-seed Won
1	3.48	27
2	4.63	23
3	4.43	24
4	4.55	27