Seeds of Victory: Big Ideas with Small Data in March Madness.

Every March, sports fans meticulously scrutinize the teams participating in the NCAA Division I Mens Basketball Championship Tournament (aka March Madness), seeking to design a perfect bracket that correctly predicts the outcomes of all tournament games. With several quintillion possible brackets, such perfection is not achieved by chance. This work proposes probability models that describe seed and bracket position performance in the tournament using a small data set, containing the results of past NCAA tournament games, and hence, these models do not rely on expert knowledge of the participating teams. Using these data, the proposed probability models estimate the likelihoods of every potential combination of seeds that can appear in each round of the tournament; statistical analysis shows that these estimates provide a close fit to the historical seed distributions in past tournaments. These models can provide a set of tools for building brackets; users wishing to rely on the models alone can generate bundles of random brackets according the estimated distributions. Comparison of the brackets generated using these models with the winners of the ESPN Bracket Challenge tournaments for the past several years is reported. (Received September 01, 2016)