

CME Interview Project

This project is property of CME Group Inc. (CME) and is used as part of their interview process. All work on this project must be your own; collaboration with anyone outside of the CME interview administrators is strictly prohibited. Distribution of the project or your results to anyone outside of the CME interview administrators is also prohibited. Any reference you use must be cited. Violations of these policies will automatically preclude you from future employment at CME.

Good luck!

June Jolly

Instructions

Please provide a write-up of your analysis in a PDF document. Make sure to state any modeling assumptions you make. If you need to write code, you may use any of the following programming languages: C, C#, C++, Java, Objective-C, Perl, Python, or Ruby. Please include your source code as part of your submission.

Introduction

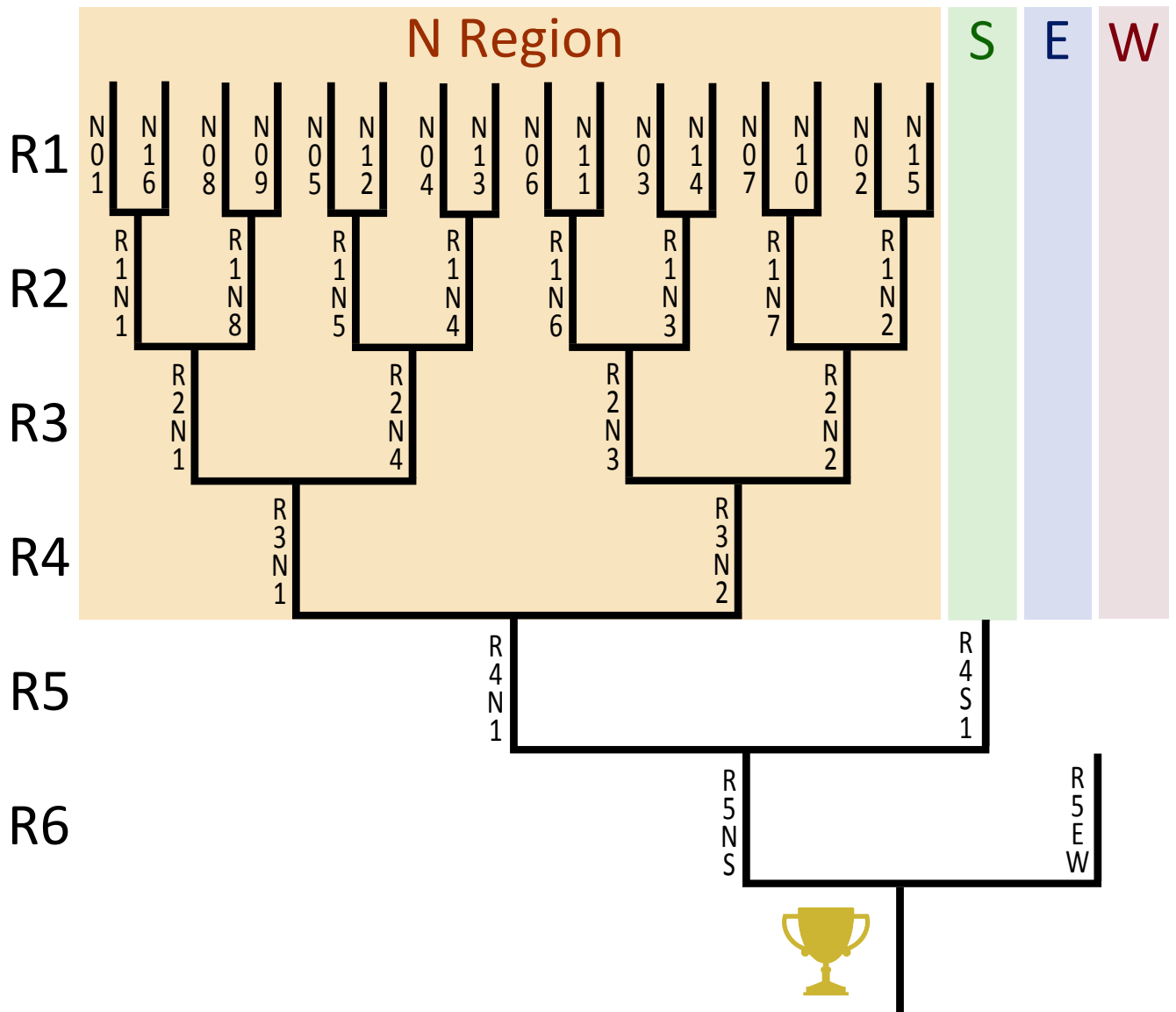
The National Funball League (NFL) runs a series of funball games each year. This time period is colloquially known as funball season. The goal of funball is to score the most points. Many cities have their own funball teams and compete in the NFL. Despite the name, funball is very serious and the games are extremely competitive.

In June after the regular season, the NFL hosts a single-elimination tournament to determine the best funball team. Win or go home! The winner gets a year's worth of bragging rights and claim to a rotating trophy.

In an attempt to make the tournament as exciting as possible, the NFL pairs stronger teams with weaker teams in the early rounds expecting that only the best teams will be left at the end of the tournament. The process of ranking teams is called seeding. Lower seed values indicate stronger teams.

For logistical reasons, the tournament is broken into 4 regions: N, S, E, and W.

The tournament is so popular that several derived games have appeared. Among the most popular is a pick'em challenge where players attempt to pick all of the winners in the tournament before any tournament games are played. In recent years, the pick'em challenge has been hosted by Sanguine Bank which offered a million dollar prize to anyone who could pick all of the correct teams.



Questions

Let p_A be the winning percentage of team A and p_B be the winning percentage of team B . Assume we use the formula 1 to estimate the probability that A wins in a matchup against B .

$$P(A \text{ wins} | A \text{ and } B \text{ play}) = \frac{p_A(1 - p_B)}{p_A(1 - p_B) + (1 - p_A)p_B} \quad (1)$$

Using this formula, which year's tournament results had the highest probability?

To answer this question, please use the attached data.

- regular_season_results.csv
- seasons.csv
- teams.csv
- tourney_results.csv
- tourney_seeds.csv
- tourney_slots.csv

Sanguine Bank has decided that the payout of the pick'em challenge is too risky. They contacted Auspicious Insurance Group (AIG) for a quote on a special insurance contract that covers the payout from the challenge. How much should AIG charge Sanguine Bank for the insurance contract? It is sufficient to explain your approach; no code or amounts are required for this question.