

# Simulating the NCAA Tournament

Joel Bracken

4/13/2020

I was very upset that my favorite sporting event, March Madness, was cancelled this year. It didn't feel right for there not to be a champion named, so I used some of the time I would be watching the games to make a basic simulator of the tournament (this could be used down the road for filling out a bracket as well). Using KenPom as a source, the data was loaded in. Only two metrics will be used from KenPom, the Adjusted Efficiency Margin and the Adjusted Tempo. The Adjusted Efficiency Margin (AdjEM) is simply the average number of points a team is expected to outscore its opponent by in 100 possessions, and the Adjusted Tempo (AdjT) is how many possessions a team gets on average in a 40 minute game.

```
data <- as.data.frame(read.csv('KenPom.csv', header=TRUE)) #load in data
kenpom <- data[, -c(1,3,4,6,7,8,9,11)] # remove extra columns
kenpom$AdjEM <- as.numeric(as.character(kenpom$AdjEM)) #adjusted efficiency margin
kenpom$AdjT <- as.numeric(as.character(kenpom$AdjT)) #adjusted tempo
```

The idea for simulating games is as follows: find the difference in AdjEM between the two teams and multiply is by the average number of possessions between the two teams divided by 200. We now have the expected points difference, but games are not always according to 'chalk', so we first calculate the probability of winning by finding the 1-the probability of being 0 or less from a normal distribution with a mean of the expected point differential and a standard deviation of 11 (suggested from KenPom). Now, to simulate the madness of March, I included a Tournamnet Randomness Factor. Let's walk through an example. If Kansas is supposed to beat Ohio State 89% of the time (on a neutral floor), we essentially then pick a random number between 0 and 100 and if that number is 89 or less, Kansas wins, if not, Ohio State wins. While this doesn't quite simulate the madness, it prevents our bracket from being too 'chalky'.

```
set.seed(2020) #for the 2020 Tourney

#function to simulate a game
simulate_game <- function(TeamA, TeamB){

  ExpectedPointDiff <- (kenpom[which(kenpom$Team==TeamA),2] -
                        kenpom[which(kenpom$Team==TeamB),2])*
    ((kenpom[which(kenpom$Team==TeamA),3]+kenpom[which(kenpom$Team==TeamB),3])/80)
  ProbTeamAWins <- 1 - pnorm(q=0, mean = ExpectedPointDiff, sd=11) # sd 11 from KenPom
  TRF <- runif(1, min = 0, max = 1) #Tourney Randomness Factor

  if (ProbTeamAWins >= TRF){
    return (TeamA)
  }
  if (ProbTeamAWins < TRF){
    return (TeamB)
  }
  else{
    return(-1)
  }
}
```

Now we must deal with the play in games. We simulate the first four games before the bracket for simplicity. Since we never got a real bracket, all tournamnet layout is product of the final 'Bracketology' on ESPN.

```

# Set up Play in Games
playin1 <- c('Boston University','Robert Morris')
playin2 <- c('Texas','Richmond')
playin3 <- c('N.C. State','UCLA')
playin4 <- c('Prairie View A&M','North Carolina Central')

#Results from Play in Games
playin1win <- simulate_game(playin1[1],playin1[2])
playin2win <- simulate_game(playin2[1],playin2[2])
playin3win <- simulate_game(playin3[1],playin3[2])
playin4win <- simulate_game(playin4[1],playin4[2])

```

Now to set up the regions from the 2020 tourney.

```

#Set up Regions

Midwest <-c('Kansas','Siena',          # 1 vs 16 seed
            'Houston','Marquette',      # 8 vs 9 seed
            'Auburn','Liberty',         # 5 vs 12 seed
            'Wisconsin','North Texas',  # 4 vs 13 seed
            'Iowa','East Tennessee St.', # 6 vs 11 seed
            'Duke','Belmont',           # 3 vs 14 seed
            'Providence','Arizona St.', # 7 vs 10 seed
            'Kentucky','North Dakota St.') # 2 vs 15 seed

East <- c('Dayton',playin1win,
          'Colorado','Florida',
          'Butler',playin2win,
          'Maryland', 'Akron',
          'Penn St.',playin3win,
          'Villanova','Hofstra',
          'West Virginia','Utah St.',
          'Florida St.','Northern Kentucky')

West <- c('Gonzaga',playin4win,
          'LSU','Oklahoma',
          'Michigan','Yale',
          'Oregon','New Mexico St.',
          'BYU','Indiana',
          'Seton Hall','Eastern Washington',
          'Arizona','Texas Tech',
          'San Diego St.','UC Irvine')

South <- c('Baylor','Winthrop',
           "Saint Mary's",'Rutgers',
           'Ohio St.','Stephen F. Austin',
           'Louisville','Vermont',
           'Virginia','Cincinnati',
           'Michigan St.','Bradley',
           'Illinois','USC',
           'Creighton','Little Rock')

#put 64 teams in a list
FieldOf64 <- c(Midwest,East,West,South)

```

Create a function for simulating an entire round.

```
#Function to simulate a round of the Tourney
#TeamsInRound is a list of teams in the current round

sim_a_round <- function(TeamsInRound){ #list of teams in round
  winners=character()
  for (i in seq(from=1, to=length(TeamsInRound)-1, by=2)){
    winners= append(winners,simulate_game(TeamsInRound[i],TeamsInRound[i+1]))
  }
  return(winners)
}
```

Create a function to simulate the entire tournament.

```
# Function to simulate the entire Tourney, input Round1 teams (FieldOf64)

sim_the_tourney <- function(Round1){
  Round2 <- sim_a_round(Round1)
  cat("The first round is complete, these teams advance to the round of 32!\n")
  print(Round2)

  Sweet16 <- sim_a_round(Round2)
  cat('\n\nThe following teams have earned a spot in the Sweet 16! \n')
  print(Sweet16)

  Elite8 <- sim_a_round(Sweet16)
  cat('\n\nThese teams kept dancing into the Elite 8! \n')
  print(Elite8)

  Final4 <- sim_a_round(Elite8)
  cat('\n\nOne shining moment for these four teams,
      who have earned their spot in the Final .\n',
      'We will see matchups of', Final4[1], ' vs ' ,Final4[2], ' and ' ,
      Final4[3], ' vs ' , Final4[4], '.\n')

  ChampGame <- sim_a_round(Final4)
  cat('\nA season to remember...the championship game will feature '
      , ChampGame[1] , ' and ' , ChampGame[2], '.\n')

  Champ <- sim_a_round(ChampGame)
  cat('\nHang the Banner! The national champion is ' , Champ, '!!!!' )
}
```

Finally, run the function! Enjoy!

```
sim_the_tourney(FieldOf64)

## The first round is complete, these teams advance to the round of 32!
## [1] "Kansas"      "Houston"      "Auburn"       "Wisconsin"
## [5] "Iowa"        "Duke"         "Arizona St."  "Kentucky"
## [9] "Dayton"      "Colorado"     "Butler"       "Maryland"
## [13] "N.C. State"  "Villanova"    "West Virginia" "Florida St."
## [17] "Gonzaga"     "LSU"          "Michigan"     "Oregon"
## [21] "Indiana"     "Seton Hall"   "Arizona"      "San Diego St."
## [25] "Baylor"      "Rutgers"      "Ohio St."     "Louisville"
```

```

## [29] "Cincinnati"      "Michigan St."  "Illinois"      "Creighton"
##
##
## The following teams have earned a spot in the Sweet 16!
## [1] "Kansas"          "Wisconsin"     "Duke"          "Kentucky"
## [5] "Colorado"        "Butler"        "Villanova"     "Florida St."
## [9] "Gonzaga"         "Oregon"        "Seton Hall"    "San Diego St."
## [13] "Baylor"          "Ohio St."      "Michigan St."  "Creighton"
##
##
## These teams kept dancing into the Elite 8!
## [1] "Kansas"          "Duke"          "Butler"        "Villanova"
## [5] "Gonzaga"         "Seton Hall"    "Baylor"        "Michigan St."
##
##
## One shining moment for these four teams,
##     who have earned their spot in the Final .
## We will see matchups of Kansas vs Villanova and Gonzaga vs Baylor .
##
## A season to remember...the championship game will feature Kansas and Baylor .
##
## Hang the Banner! The national champion is Baylor !!!!

```