

CAROLINAS SPORTS ANALYTICS MEETING 2014

PRESENTATION ABSTRACTS

FBS Conference Realignment

Mikaela Cashman, Coe College

Coauthors: Brent Davis, Jeremy Santiago

We develop an algorithm to place the set of 125 NCAA Division 1 Football Bowl Subdivision (FBS) teams into new conferences. This algorithm focuses on a set of parameters for identifying a conference realignment which defines the equity of football. Our algorithm maximizes the equity to produce “good football”. We then wrote a program to test combinations of conferences then pick the highest ranking conference alignment.

Designing a college conference tournament

Gavin Cross, Coe College

There are many ways to design a conference tournament regardless of the sport being contested. Designs can either reward teams’ good seasons or just the current ‘hot’ teams. Different designs will be discussed and analyzed. These designs will include both conference tournament formats that are currently in use and those that are not in use. Analysis will include: double vs. single elimination methods, ranking methods, & seasonal record methods.

The Gender Gap and the Two-Hour Mark in Major Marathon Races

Kirstie Doehler, Elon University

Coauthor: Seong-Tae Kim

In sports analytics and physiology, researchers have addressed if and when a two-hour marathon will happen as well as whether the gender gap in marathon winning records will decrease. Our study aims to reanalyze these two issues with extended data from the major marathons and more sound statistical methods beyond simple linear regression. We investigate men’s and women’s winning records from the five major world marathons: Boston, New York City, Chicago, London and Berlin. The historical records of these races were substantially decelerated to a plateau between the mid-1970s and mid-1980s. When building the prediction models of the two-hour record and the gender gap, this nonlinear feature is considered using various statistical methods such as linear regression with change points, nonlinear regression, and nonparametric regression. We present analytical methods and results which are compared to previous outcomes published in the literature related to marathon records.

Random Matchups and NBA Ratings

Dante Durrman, Furman University

We propose a simple ranking system for NBA teams loosely based on a bootstrapping paradigm. First we generate a team-by-team matrix of probabilities based on real-time statistics and then we simulate outcomes of randomly generated knockout tournaments by resampling this data.

College Sports Rankings via HodgeRank

Ryan Gutschick, Wake Forest University

In this talk, we will examine the foundation and principles behind HodgeRank, a ranking method that takes pairwise comparison data and establishes a relative ranking and rating of each data point. In particular, we will use the scores of 2013 college football and 2014 college basketball match-ups as our pairwise comparison dataset (a very natural system for such comparisons) to investigate the theories that underlie HodgeRank including the Hodge Decomposition Theorem and its use of gradient, harmonic, and flux-less (or curly) subspaces. Additionally, we will discuss how HodgeRank can be used to determine the reliability of our rankings as well as determine relatively how much “parity” actually exists in collegiate sports today.

Identifying Influential Upsets in College Football

Kevin Hutson, Furman University

Networks model complex systems from a wide range of applications, including contests between teams during a sport season. One of the well-studied aspects of a network is the concept of the most influential or powerful node(team) in the network. The edges (games) of the network can also exert great influence in such measures of node power. In this talk, we propose methods that identify influential edges in a network and their impact on a ranking of nodes as it applies to a college football season.

The Effects of League Parity and Season Length on NFL Team Success

Jai Kedia, College of Wooster

Coauthors: Sebastian Weber, R. Drew Pasteur, and Robert D. Wooster

We attempt to determine the effects of varying season length, league parity, and home-field advantage in the NFL. In a fictionalized league with a structure mirroring that of the NFL, we assign team strengths by sampling from a normal distribution, then simulate a season, including playoffs. By repeating over many seasons, we can estimate the probabilities of the teams of various strengths making the playoffs, winning the Super Bowl, etc.

Can Coase govern the Greenbay Packers?

Manish Khadka, Coe College

Coauthor: Jay Chen

My research in prior summer was been on the corporate control modeled on the distinct organizational structure of the Green Bay Packers, the only corporate owned franchise in the NFL. My study focuses on whether the presence of the Packers' diverse ownership base will lead to severe agency costs (costs incurred when the interests of shareholders and management are not aligned). To test the hypothesis, I look at measures that suggest whether Packers are better or worse than the average NFL teams over the past 4 decades. The results are contrary to the agency theory prediction. I further study whether the Coase Theorem can help explain why the Packers perform in such a distinctive pattern. I conclude that the implicit transfer of ownership from shareholders to management is the key success factor for the Packers, a result consistent with the Coase Theorem.

Predicting Atlantic Sun Volleyball Tournament Outcomes

Lizzie Knapper, Mercer University

Coauthor: Hope McIlwain

Predicting tournament brackets is a favorite activity for many sports fans. Although there is no flawless way for prediction, we use mathematical models to predict the outcome games. In this project, we use two matrix analysis methods in our search for an accurate prediction method for the outcomes of games in Women's Indoor Volleyball in the Atlantic Sun Conference. In the Massey Matrix Method, the ratings depend on the point differentials, or more generally, in the differences in a given statistic. Keener's Method ranks teams, not on wins and losses, but on the values of non-negative statistics that result from the competition. Keener's method measures the absolute strength of the team based on the relative strength of the team, i.e. how strong a team is compared to its competitors. We use three different statistics when running these methods: kills per match, match scores, and total points. So we have a total of six different prediction methods. Throughout the season, we recalculate weekly rankings for Atlantic Sun Conference matches. From the rankings, we predicted the outcome for the matches in the following week. We then calculate the percentage of accurately predicted games, called the forecast accuracy. At the end of the season, we consider the overall number of accurately predicted games to determine the best prediction method.

Picks for Kicks (How FIFA's Structure Complicates Ranking National Soccer Teams)

Mark Kozek, Whittier College

Coauthors: Tim Chartier and Michael Mossinghoff

In the first of our two talks on ranking national soccer teams, we provide a background on the FIFA World Cup. We give a brief history. We describe the competitions that lead up to the World Cup and the World Cup's playoff structure. And we highlight certain peculiarities in these competitions that make ranking national soccer teams (and thus predicting the World Cup) more difficult than March Madness or the BCS.

Doing Sports Analytics Research: Lessons from a Beginner

Hope McIlwain, Mercer University

Beginning research in a new area can be a challenge. Several years ago, I switched from studying graph theory to studying sports analytics. Along the way, I learned many lessons about how to do sports analytics as research through making mistakes, learning from others, and doing several different types of projects. In this talk, I will discuss those lessons.

Putting Putting In Its Proper Place

Roland Minton, Roanoke College

The statistical analysis of golf has been revolutionized by ShotLink technology, by which the location of every shot on the PGA Tour is determined to the inch. The new level of precision has allowed Strokes Gained to be developed, where Strokes Gained measures the value of a stroke by comparing the golfer's expected score before and after the stroke. By comparing Strokes Gained Putting to Strokes Gained Driving or Strokes Gained for other skills, the importance of each skill in golf can be estimated. An old standard bit of golfing folklore is that putting is everything ("drive for show and putt for dough") but the data tell a different story. Time series of PGA golfers' Strokes Gained in tournaments are analyzed in a variety of ways to try to get a handle on the importance of putting and other skills on the PGA Tour.

Implicitly Defined MLB Batter and Pitcher Statistics

Darin Mohr, Georgia College

Classic statistics in Major League Baseball (MLB) such as batting average for batters and earned run average for pitchers do not incorporate the skill of the opposing pitchers and batters respectively. We present new metrics for batter and pitcher performance that incorporate the relative skill of the opponents and apply these metrics to evaluate the 2013 MLB season.

FIFA Foe Fun!

Mike Mossinghoff, Davidson College

Coauthors: Tim Chartier and Mark Kozek

Who will win the World Cup this summer? Do you smell the blood of an Englishman? Will defending champion Spain come out on top? Host nation Brazil? Another perennial power, or a dark horse? Who will advance into the knockout round? Which one is the Group of Death? After reviewing FIFA's official ranking method, we investigate these questions by applying a number of ranking methods. Our methods take into account the result of every match between national teams held since the last World Cup. We also discuss the parameters selected in these systems to account for certain peculiarities in FIFA competition.

Evaluating NFL Field Goal Kicker Performance by a Value Added Metric

Drew Pasteur, College of Wooster

Coauthor: Kyle Cunningham-Rhoads

In American football, the standard measure of placekicker quality is field goal percentage, the fraction of attempts successfully converted. Due to variance in distance and other conditions (weather, altitude, etc.), we find this to be an inadequate way to rate kickers. Using three seasons of NFL data, we build a logistic regression model for the success probability of a given attempt. This allows us to compare a kicker's aggregate performance to that of an inexpensive "replacement" player, over the same set of kicks, leading to value-added measures similar to baseball's WAR statistic. We then compare our list of the best kickers to those of the market (using player salaries) and popular opinion (Pro Bowl and All-Pro selections). Finally, investigating the common belief that kickers are largely interchangeable, we examine the degree to which kicker performance is maintained from one season to the next.

Understanding Intracontinental Trade in Europe Using Rankings

Sophia Rivera

Coauthor: Hope McIlwain

The objective of this research is to further understand the power of intracontinental trade in European. We are performing the Colley, Massey, and Keener Method on the import and export data from eleven different countries in the Europe. These pair-wise comparisons will be used to calculate export wins and losses. We define an export win for country A if country A exports more to country B than country B exports to country A. An export loss is similarly defined, but country A sends fewer exports to country B than country B sends to A. Using these three different methods to create a ranking we will better understand the status of each economy compared to the others within Europe, we hope to find which methods are more promising in understanding the current trade interactions between selected countries.

A Monte Carlo Simulation Approach to Determine the Greatest Post-Season Players

Abigail Stryker, Muhlenberg College

Coauthors: Erica Wenzel, Michael Huber, and Ashlie Christian

Until 1969 the "post-season" consisted of the World Series. Now the post-season consists of a multigame tournament of five teams from each league. Therefore, a post-season that originally had at most 7 games may now consist of at most 20 games, leading to an inequity in post-season player comparison. In this paper, we research and analyze actual data of players and use computer simulations to predict post-season home run totals if players like Ruth and Mantle had played in more post-season games. By doing this, we can determine the most successful players in the post-season. To check the validity of this simulation, we use the scheme on current players and compare to known data.

The Impact of the Number of Players on Ice in NHL Games

Marty Thomas, Georgia Gwinnett College

Research has shown the dislike that American fans have for games that end in ties. Many professional sports organizations have made rule changes to either decrease the likelihood or eliminate altogether games that end in a tie. Each team in National Hockey League (NHL) games typically has 5 skaters on ice (3 forwards and 2 defensemen). In 1998, the NHL changed the 5 minute overtime period to one where each team has 4 skaters on ice (2 forwards and 2 defensemen) in an attempt to create more open play. The NHL is presently considering an additional 5 minute overtime of 3 on 3 skaters for games that are still tied after the 5 minute 4 on 4 skaters overtime period. This research will make use of the NHL's real-time play-by-play database to determine the rate per minute of significant events (face offs, hits, shots on goal, etc.) in terms of the number of skaters on ice. These results will allow for predictions regarding the rate at which significant events occur and the likelihood of a goal within a 5 minute segment of 3 on 3 skaters on ice.