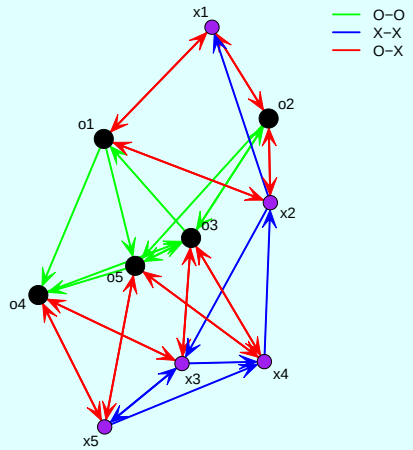
Notes from the Court: Some Practical Ideas for Doing Analytics

Aaron Danielson

UCLA Department of Statistics

February 22, 2018



Overview

1. Who is this person?
2. What do we mean by "Analytics?"
3. Analysis Tips to Consider
4. Sports Statistics: Culture, Projects and Data Management

Introduction

- PhD candidate in the statistics department.
- Actually came to UCLA as a PhD student in political science; Switched to statistics after my adviser left UCLA.
- Study network methods, sampling and model selection.
- Interested in applications in politics, culture and sports.
- Hold degrees in economics, public policy and philosophy. (Ahhh ... student debt!)
- Taught English in a small town near Nagasaki, Japan.
- Worked for the LA Lakers as the primary data analyst for 3 years.
- Club athlete in college (water polo).
- Husband and father.

On Analytics

Main Ideas

- Analytics could be defined as the application of scientific techniques to produce results communicated to decision-makers.
- Projects work well when the goals are clearly understood by the analyst and the final consumer of the research.
- The method(s) chosen should depend upon the goal of the research, type of data available and time horizon.
- Role of an analyst can vary greatly between institutions.
  - Does your supervisor know much about statistics?
  - Do you work on a specific problem or a variety of projects?
  - Is the data difficult to access, store and manipulate?

- Designing Experiments and Surveys: examples include randomized trials, conjoint analysis, opinion research.
- Collecting Data: survey implementation, web scraping, field work, case study.
- Database Management: Design system to efficiently store, query, explore, and manipulate data.
- Creating Visualizations and Data Tables: dashboards, reports, apps, exploratory analysis.

- Building Models: estimate return on some attribute, power rankings for team sports, find expected time to injury for a player.
- Hypothesis Testing (Answer questions posed by decision-makers):
- Forecasting: predict some quantity of interest each quarter, compute counterfactual circumstances, Draft preparation models.
- Strategy (Decision Theory/Game Theory/Statistical Mechanics): when to foul, how far is too far for threes, geometry and play success.
- Communicate analysis to various audiences in written and oral format in meetings and presentations.

On Analytics

My Personal Bias

Play Pause

- Subject Matter Knowledge Matters
  - Must know what to study.
  - Cannot evaluate assumptions without some degree of knowledge.
  - Do you have competing theories to explain your data?
- Focus on exploratory analysis of complex data sets
  - Do simple things and visualize the results.
  - Use dimension reduction techniques.
- Use causal methods when appropriate and interpret them as such.
- Do not rely on the same set of procedures: If a computer can do your job, you may not have it for very long.
- *Parsimony* and *interpretability* are desirable, but may be unrealistic.
- Proudly report null results when they arise.

On Analytics

My Personal Bias

- Preprocess complex data to aid analysis.
- Implement several methods to check robustness of results.
- Use model selection techniques (preferably Bayesian) when no single data-generating mechanism is evident.
- Take responsibility for your analysis: You ARE partially responsible for actions taken based upon your work.
- Follow principles of data ethics.
- Data does NOT *speak for itself*: the world has never told us anything about how it works.
- Use qualitative information. The quantitative/qualitative distinction should not limit you.

TNT on Analytics in Sports

- Actual quote from a meeting, "Look, none of us have much faith in academic research, but ..."
- Analysts need to work well with non-technical staff to generate and implement ideas.
- Statistics is one of many ways to learn from experience. Be humble - your presence may threaten others.
- Is scouting a form of analytics? (Question asked to crowd at MIT Sloan Sports Analytics Conference by Mike Zarren of the Celtics).

Statistics in Sports

What do Analytics Staffers do?

- Quantitatively scout opponents
- Study the draft
- Construct and evaluate trades
- Analyze lineups
- Suggest roster changes
- Assess risk of injury to players
- Model league standings
- Adjusting strategy
- Collect data from practices
- Design player development plans
- And much more ...

Statistics in Sports

Managing and Using SportVu Data

- Spatio-temporal information from sports contests: soccer, hockey and basketball.
- Six cameras in the rafters.
- Contains id codes for players, location (x, y coordinates) of all players and the location of the ball (x, y, z) at intervals of .04 seconds (25 times per second).
- Camera records data continuously so analysts must be careful not to include observations from timeouts or intermission.
- Length of data varies (overtime, lots of free throws, player injuries, delay in game), but typically 200,000 observations per game.
- Could use SQL within R to develop with the data. In general, it was easier to load a single game or quarter at a time.
- Competitors such as Second Spectrum are working on improving the camera-generated data (shorter time intervals, z coordinate for players).

- Data arrives early each morning (4 AM) in XML format (JSON also common).
- An automated script loads the new data into Hadoop.
- A processing script loads the new data and creates a set of processed files in table form.
- These processed files are aggregated at the possession, game or season level.
- Examples include: identifying drives, computing workloads (acceleration over time), analyzing passes, computing geometric quantities such as the area of a convex hull, etc ...
- An analysis script runs on the processed data and produces a written report.

- Start doing analytics by publishing results on a blog.
- Attend meetings such as the MIT Sloan conference and the New England Symposium on Statistics in Sports (NESSIS).
- Volunteer to work for a high school or college team.
- Apply for a job at a sports analytics or media company.
- Network, network, network.
- The truth is that most jobs with teams are very hard to get as teams tend to hire people they know over outsiders.
- There is a need for theory to make better use of the data - this is the focus of some of my work.

Statistics in Sports

Emerging Opportunities

- Professional, College, High School and AAU teams.
- Sports Analytics and Finance.
- Sports Analytics and the Media.
- Academic Positions.

Thank you!