Sports Analytics Overview

Sports Analytics Overview

- Data infrastructure and management
  - Data obtainment, Big data techniques
- Programming skills
  - Data manipulation and analytic execution
- Methods
  - Summary analysis: descriptive and visualization
  - Estimation and inference
    - Regression analysis: frequentist and Bayesian
    - Data Mining techniques
  - Model underlying process
- Applications (Production of winning)
  - Roster design
  - Player valuations
  - In-game strategies
  - Predictions

Methods overview

- Most SA applications are complicated
    - Continual influence from multiple factors
- Summary analysis is not sufficient to make accurate conclusions
- Require modeling of the processes
- Regression analysis is a popular method for modeling processes

Statistical significance and estimation

Use mean and standard deviations to determine confidence mean difference is from randomness.

Example: goal scored per game

	Team A	Team B
Mean	2.5	3.1
Standard deviation	(0.75)	(0.55)

Regression analysis

Impose causality to estimate (linear) magnitude and significance levels of multiple dependent variables and an independent variable.

Example: Goals throughout the progression of a game

Bayes Rule

Formula for calculating conditional probabilities

$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x \mid y)p(y) + p(x \mid \bar{y})p(\bar{y})}$$

Bayes Rule example

Calculate the probability of a home team win given the score differential is the home team trailing by two goals.

$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x \mid y)p(y) + p(x \mid \bar{y})p(\bar{y})}$$

Prior probability of a home team win: 0.60

Inverse conditional (win): 0.12

Inverse conditional (loss): 0.30

Bayes Rule example

Calculate the probability of a home team win given the score differential is the home team leading by two goals.

$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x \mid y)p(y) + p(x \mid \bar{y})p(\bar{y})}$$

$$p(y \mid x) = \frac{0.12 \times 0.60}{0.12 \times 0.60 + 0.30 \times 0.40}$$

$$p(y \mid x) = 0.39$$