Stat prof assesses the Game day data

Jay Emerson
Assistant Professor of Statistics

The first game: November 13, 1875 at Hamilton Field, New Haven. The winner: Harvard, 4-0, according to different sources. According to IvyLeagueSports.com, the score was Harvard 4-1-0 over Yale 0-0-0. According to the Yale Record (November 2001), Harvard scored four touchdowns and four field goals.

Confused? I am. The problem was not as simple as I expected. A statistician must confront data with skepticism. Mistakes are sometimes obvious (data from the Yale Daily News contained an error in 2005, for example) and are easily fixed. Other times, answers elude even the most skilled Googler.

I learned that the 1875 game was essentially rugby. Terms like "goal," "touchdown," "try," "kick," and "brutality" relate to Yale's victory in 1878, for example, 1-7-0 to 0-0-13. FootballResearch.com explained that the Rugby Union Rules determined the winner of a match by the majority of "goals," but excluding "touchdowns." In the 1800's, Yale and Columbia bickered with Harvard and Princeton about the rules, eventually agreeing that a goal would be equal to four touchdowns. I still can't explain the 1878 score.

The rules of football evolved under the watchful eye of the "Father of American Football," Yalie Walter Camp (Class of 1880). By 1920, the rules appear to have settled into their current form, with touchdowns contributing six points (plus an extra point, or point after attempt, PAT), and field goals three points.

Since 1920, Yale has won 41 games to Harvard's 40. There have been three ties, and a lapse in play during World War II. Yale has outscored Harvard, 1359-1279 (barring further data errors). There is no apparent home field advantage (Yale is 21-21-1 in New Haven – 21 wins and losses with 1 tie – while Harvard is 19-20-2 in Cambridge).

In an effort to make an honest prediction for the upcoming Game, I first turned to the bookies. The gambling industry can be a wonderful source of data for predicting the outcomes of sports events (a statistical observation, not an endorsement of gambling). It is an efficient market: the predictions reflect all the information used by bettors. The losers essentially pay the winners, and the bookies happily walk away with the "juice" – fees charged for placing the bets.

Unfortunately, I was unable to find a Web site publishing odds for The Game. My usual source for data, GoldSheet.com, didn't have past gambling lines for Games, either. I conclude that there isn't enough demand to justify online gambling on The Game.

Next, I decided to display the data since 1920. The solid segments indicate Yale victories (with the Y marking Yale's winning score, and the bottom of the tail indicating Harvard's losing score). Similarly, the top of a dotted segment indicates Harvard's score in a Crimson victory (with Yale's losing score marked by the Y). Three T's indicate ties. There weren't many high-scoring games prior to World War II, and shutouts (with one team failing to score) were rare in the last few decades.

I also compared Yale and Harvard's margins of victory (or loss) against the other six teams in the league. The standout opponent is Penn, a result that could bode well for Yale in the 123rd Game this Saturday.

Finally, I collected the Ivy League results of the past 50 seasons, 1956-2005. Yale and Harvard each play six games against other Ivy opponents prior to The Game, so I calculated the number of pre-Game wins for each school in each season. I used the difference between Yale's Ivy wins and Harvard's Ivy wins (in each year) to predict the subsequent outcome of The Game. These differences range from -5 (when, for example, Yale had 1 win to Harvard's 6 prior to The Game) to +4 (in 1956 Yale had 6 Ivy wins to Harvard's 2, for example). The win difference was a statistically significant predictor of the outcome of The Game (p=0.04). I explored the possibility of a home field advantage, but found that it was not a statistically significant predictor (p=0.79).

This year, the difference in Ivy wins is +1 (Yale has 5 wins to Harvard's 4). My model predicts the probability Yale wins The Game on Saturday, November 18, 2006, to be 53.3%. Game on.

The Game, 1920–2005

Legend:
— Yale Victory
⋯ Harvard Victory
Y — Yale's Score in The Game
T — Tie Game

World War II: no Games during 1943 and 1944

Points Scored (y-axis); Year (x-axis): 1920, 1940, 1960, 1980, 2000

Pre-Game Ivy League Games Won

Yale's Ivy League wins in each season prior to The Game: 6 3 0 4 6 3 1 3 4 3 3 6 6 5 5 3 4 4 6 5 5 5 3 6 5 5 3 1 4 2 2 5 2 6 4 3 2 1 2 2 1 1 4 5 3 1 4 4 3 4

Harvard's Ivy League wins in each season prior to The Game: 2 2 2 3 4 5 4 4 4 2 5 4 5 2 4 3 3 5 5 5 4 4 2 2 4 4 4 5 5 2 5 2 4 3 4 2 1 2 0 1 6 3 3 4 6 5 3 6 4