Rod Doe

StatR101

Final Project Musings

# Summary

The statistic I would like to analyze is the total goal differential (goals scored – goals allowed) achieved by the Seattle Sounders during the Major League Soccer regular season. The factors in this analysis are the players in the lineup at every minute in the match. In particular, I want to understand the contributions and interactions by a particular player, Mauro Rosales.

By consensus, the team plays better when Rosales is on the field. He typically plays in an outside midfield position. He rarely scores goals, but does get assists. Quite conveniently, he was injured for an extended period, during which the team struggled.

# Source of Data

I had hoped to find juicy RSS feed from which I could extract soccer data. Such feeds exist; they appear to be related to gambling interests, and the data are not freely available. These do not appear to be organizations with which I want to share my credit card number. However, a free, albeit less automatable option exists – the match reports on the Sounders web site. This report is excerpted from <http://www.soundersfc.com/matchday/matches/2012/regular/mls-14-at-montreal-impact.aspx>, which reports on a 4-1 loss to the Montreal Impact.

**Scoring**   
**Sounders FC –**Eddie Johnson (Fredy Montero) 61.  
**Impact –** Felipe Martins (Patrice Bernier) 18; Justin Mapp (Patrice Bernier) 51; Andrew Wenger (Patrice Bernier) 58; Lamar Neagle 87.

**Lineups**   
**Sounders FC –** Bryan Meredith, Zach Scott, Jeff Parke, Jhon Kennedy Hurtado, Marc Burch, Mauro Rosales (Cordell Cato 71), Andy Rose, Brad Evans, Alvaro Fernandez (Alex Caskey 62), Fredy Montero, Eddie Johnson (Servando Carrasco 82).  
**Impact –** Donovan Ricketts, Zarek Valentin, Matteo Ferrari (Shavar Thomas 32), Nelson Rivas, Jeb Brovsky, Davy Arnaud, Patrice Bernier (Bernardo Corradi 77), Felipe Martins, Collen Warner, Justin Mapp (Lamar Neagle 62), Andrew Wenger.

From these reports, I can extract:

* The players on the field during each minute of the match.
* The minute in which each goal was scored
* The minute in which each goal was allowed.

For example, the Sounders scored in the 61st minute. Montreal scored in the 18th, 51st, 58th, and 87th minute. Zach Scott played 90 minutes, while Mauro Rosales was substituted in the 71st.

The Sounders will play 34 regular season matches. As such, my data will consist of 90 \* 34 rows, a column for total goals scored for each minute, a column for total goals allowed for each minute, and a column for each player who entered a match (approximately 25). A player would get a T or F for each minute of every match played. These data would be fairly easy to construct with Excel.

The goal differential would be modeled as a function of the lineup. This is a concern (analysis might melt my laptop), since with interactions of different player combinations, the model could look something like this:

GoalDifferential ~ Meredith\* Scott\*Parke\*Hurtado\*Burch\* Rosales \* Cato \*Rose\*Evans\*Fernandez\* Caskey \* Montero\* Johnson \* Carrasco

# Questions

* This seems like it could produce hopelessly complex results. Do you agree?
* Would it be better just to replace each \* in the model with a + and not look for interactions?
  + It seems that interactions are an important part of this analysis, i.e., interactions would identify the ideal lineup.