# StatR 101: Fall 2012

Final Project

Rod Doe

TBD: Wedneday, November 28, 2012

# Home Field Advantage in the Major League Soccer 2012 Season

This project quantifies home field advantage in Major League Soccer (MLS). In the 2012 season, of the 338 matches played, the win-draw-loss record for the home side was 172 wins, 82 draws, and 84 losses. Locally, the Seattle Sounders had 11 wins, 2 draws, and 4 losses on at home, and had 4 wins, 9 draws, and 4 losses on the road.

These data are available from [www.goal.com](http://www.goal.com). I harvested data for all MLS teams from this web site using a “screen scraper” script I wrote in Windows PowerShell. This script appears in Appendix TBD.

The data from this site include data from matches between MLS and non-MLS teams. Results from non-MLS competition were excluded.

Data on MLS attendance are available, oddly enough, from [www.mlsattendance.com](http://www.mlsattendance.com). Data on MLS team salaries are available from Wikipedia.

Two MLS teams play in the same stadium. For this reason, data from matches between these teams (LA Galaxy and Chivas USA) were excluded from home field advantage analysis.

To obtain the distances traveled by visiting teams it was necessary to:

* Obtain the latitude and longitude of the each city’s airport.
* Calculate the great circle distance between these cities ([nautical miles]).

## The Significance of Home Field Advantage

My null hypothesis is that home field advantage is a myth, and that one should expect the results of the matches played to be equally split between home team wins, draws, and home team losses.

H0: There was no advantage to play on home field in the 2012 MLS soccer season.

Count of home wins = count of draws = count of home losses.

H1: There was an advantage to play on home field in the 2012 MLS soccer season.

Count of home wins > count of draws and count of home wins > count of home losses.

In the MLS 2012 season, 172 matches were won by the home side, 82 matches were draws, and the home side lost 84 times. A chi-squared analysis of these proportions reveals that this result is quite unlikely to occur by chance.

mls<-read.csv(file="MLSWinDrawLose2012.csv")

wdlCount <- matrix(c( sum(mls$HomeWin[mls$NauticalMiles > 0]),

sum(mls$Draw[mls$NauticalMiles > 0]),

sum(mls$HomeLoss[mls$NauticalMiles > 0])), nrow=1)

chisq.test(wdlCount)

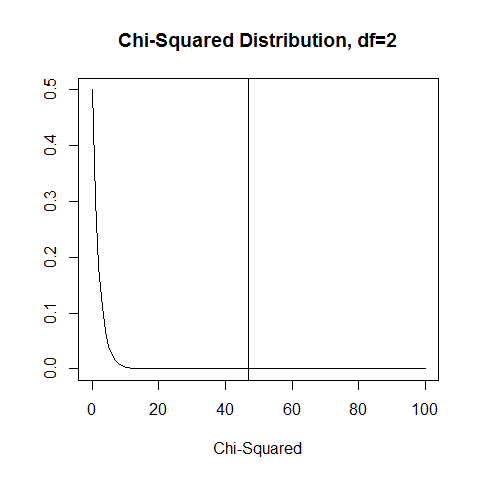
Chi-squared test for given probabilities

data: wdlCount

X-squared = 46.8876, df = 2, p-value = 6.584e-11

The graphical presentation of these data requires a function to get the distribution of a chi-squared distribution with two degrees of freedom.

dwdl <- function(x) {dchisq(x, df=2)} #dwdl is chisq distribution of win-lose-draw



The area under the curve indicates a strong effect:

> integrate(dwdl, 0, 46.9)

1 with absolute error < 3.1e-06

So, the data establish that home field advantage is highly statistically significant given the miniscule alpha value indicated by the chi-squared result. I reject my null hypothesis.

## Home Team Goal Differential

A simple statistic to measure the home team advantage is home team goal differential. This value is simply the difference between the home team score and the visitor team score. Here are the 2012 results for each team.

|  |  |
| --- | --- |
| **Team** | **Home Goal Differential** |
| Chicago Fire | 0.44444444 |
| Chivas USA | -1.20000000 |
| Colorado Rapids | 0.58823529 |
| Columbus Crew | 0.41176471 |
| D.C. United | 1.05263158 |
| FC Dallas | 0.29411765 |
| Houston Dynamo | 1.21052632 |
| Los Angeles Galaxy | 0.70000000 |
| Montreal Impact | 0.66666667 |
| New England Revolution | 0.47058824 |
| New York Red Bulls | 0.83333333 |
| Philadelphia Union | -0.05555556 |
| Portland Timbers | 0.17647059 |
| Real Salt Lake | 0.61111111 |
| San Jose Earthquakes | 1.05555556 |
| Seattle Sounders FC | 0.89473684 |
| Sporting Kansas City | 0.57894737 |
| Toronto FC | -0.58823529 |
| Vancouver Whitecaps | 0.47058824 |

Summarized…

|  |  |
| --- | --- |
| **Category** | **Home Team Goal Differential** |
| MLS Overall | 0.47 |
| Eastern Conference | 0.51 |
| Western Conference | 0.44 |

Give the data available, are there possible correlation to the home team goal differential. These factors are explored:

* Distance travelled by the visitor.
* Time zones traversed by the visitor.
* Home team stadium size.
* Home team stadium utilization.
* Team salary differential.

Other questions arise:

* Is there any observed difference between the effects of travel within the same time zone as opposed to across time zones?
* Does Seattle, in one corner of the country, enjoy a better home field advantage due to travel as opposed to a more centrally located team such as the Chicago Fire?
* Is the home field advantage less among teams in the same time zone? For example, in the MLS Eastern conference, D.C United, New York Red Bulls, Columbus Crew, Montreal Impact, Philadelphia Union, New England Revolution, and Toronto FC are all in the same time zone.

In addition, I have collected data available from [www.soundersfc.com](http://www.soundersfc.com) on a particular player of interest, Mauro Rosales. Rosales is a player who doesn’t score a lot of goals, but my qualitative observation is that his presence is a catalyst for improved team performance. I would like to quantify this. He is a somewhat frail player, injured about 1/3 of the time. These absences present an opportunity to compare team performance with and without him. This is just a curiosity – I would like to say with N percent confidence that the team plays better with him, if only to quantify how much I should panic whenever he is injured.