

Aggressiveness

Andrew Lim

2017-07-08

```
suppressMessages(library(tidyverse))

data <- suppressMessages(read_csv("../data/AllActionsperMatchUSAMajorLeagueSoccer2016.csv"))
match_data <- suppressMessages(read_csv("../data/TeamDataMLS2016.csv"))

## Warning: Duplicated column names deduplicated: 'goals' => 'goals_1' [49]

match_data <- match_data %>%
  rename(match_result = Result)

# Joining the two tables:
data <- data %>%
  left_join(match_data, by=c("Match" = "Matchid", `Team in possession` = "Team"))

aggressive_df <- data %>%
  mutate(next_team = lead(team)) %>%
  mutate(turn_over = team != next_team) %>%
  mutate(aggressive_turnover = LocX > 50 & turn_over == TRUE) %>%
  mutate(aggressive_turnover_by_team = aggressive_turnover == TRUE & `Team in possession` == next_team)
  select(`Team in possession`, next_team, aggressive_turnover, aggressive_turnover_by_team, everything())
  group_by(Match, `Team in possession`) %>%
  summarize(num_aggressive_turnovers = sum(aggressive_turnover),
            match_result = unique(match_result),
            HomeAway = unique(HomeAway)) %>%
  mutate(is_home = ifelse(HomeAway == "T", TRUE, FALSE)) %>%
  mutate(is_win = match_result == "W")

aggressive_df

## Source: local data frame [680 x 7]
## Groups: Match [340]
##
## # A tibble: 680 x 7
##   Match `Team in possession` num_aggressive_turnovers match_result
##   <int> <chr> <int> <chr>
## 1 23579 New York Red Bulls 217 V
## 2 23579 Toronto FC 136 W
## 3 23580 Orlando City SC 135 G
## 4 23580 Real Salt Lake 110 G
## 5 23581 Chicago Fire 131 V
## 6 23581 New York City FC 163 W
## 7 23582 FC Dallas 171 W
## 8 23582 Philadelphia Union 147 V
## 9 23583 Houston Dynamo 151 G
## 10 23583 New England Revolution 162 G
## # ... with 670 more rows, and 3 more variables: HomeAway <chr>,
## # is_home <lgl>, is_win <lgl>
```

```
# The following code was obtained from: https://sebastiansauer.github.io/convert_logit2prob/
logit2prob <- function(logit){
  odds <- exp(logit)
  prob <- odds / (1 + odds)
  return(prob)
}
```

```
# Teams tend to play more aggressive when they are at home:
model <- lm(num_aggressive_turnovers ~ is_home, data=aggressive_df)
summary(model)
```

```
##
## Call:
## lm(formula = num_aggressive_turnovers ~ is_home, data = aggressive_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -57.147 -14.538  -1.538   13.462   76.462
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   130.147      1.102  118.126 < 2e-16 ***
## is_homeTRUE    10.391      1.558   6.669 5.35e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.32 on 678 degrees of freedom
## Multiple R-squared:  0.06156,    Adjusted R-squared:  0.06018
## F-statistic: 44.48 on 1 and 678 DF,  p-value: 5.346e-11
```

```
# Teams should be more aggressive when they are away in order to best improve their chances of winning:
model2 <- glm(is_win ~ num_aggressive_turnovers + is_home + is_home*num_aggressive_turnovers, data=aggressive_df)
summary(model2)
```

```
##
## Call:
## glm(formula = is_win ~ num_aggressive_turnovers + is_home + is_home *
##      num_aggressive_turnovers, data = aggressive_df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7817  -0.2872  -0.1567   0.4177   0.9751
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.651766   0.152234   4.281 2.13e-05
## num_aggressive_turnovers -0.003562   0.001156  -3.082  0.00214
## is_homeTRUE       0.883470   0.225062   3.925 9.54e-05
## num_aggressive_turnovers:is_homeTRUE -0.003826   0.001643  -2.329  0.02016
##
## (Intercept)          ***
## num_aggressive_turnovers      **
## is_homeTRUE           ***
## num_aggressive_turnovers:is_homeTRUE *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1887578)
##
##      Null deviance: 153.16  on 679  degrees of freedom
## Residual deviance: 127.60  on 676  degrees of freedom
## AIC: 801.99
##
## Number of Fisher Scoring iterations: 2
```

$$P(win) = \frac{1}{1 + e^{-x}}$$

Where:

$x = 0.65 - 0.0036 * (\text{number of aggressive turnovers}) + 0.883 * (\text{is at home}) - 0.033 * (\text{number of aggressive turnovers} * \text{is at home})$

```
1/(1 + exp(-0.65))
```

```
## [1] 0.6570105
```

The above is the chances of winning under the following conditions: * Away * Zero aggressive turn overs

```
1/(1+exp(-(0.65 + 0.883)))
```

```
## [1] 0.8224448
```

The above is the chances of winning under the following conditions: * Home * Zero aggressive turn overs