STA210 Final Project

Aryan Mathur, Molly Honecker

Introduction

```
library(tidyverse)
library(tidymodels)
library(lme4)
library(lubridate)

baseball <- read_csv("baseball.csv")
colnames(baseball) <- gsub(" ", "_", colnames(baseball))
colnames(baseball) <- tolower(colnames(baseball))

baseball$month <- month(as.Date(baseball$date))
baseball$score_diff <- baseball$home_team_score - baseball$visiting_team_score
baseball$home_win <- ifelse(baseball$home_team_score > baseball$visiting_team_score, 1, 0)

m1 <- glm(home_win ~ attendance + time_of_game_minutes + day_night_indicator + day_of_week
          data = baseball,
          family = 'binomial')

summary(m1)
```

```
Call:
glm(formula = home_win ~ attendance + time_of_game_minutes +
    day_night_indicator + day_of_week + home_team_game_number +
    visiting_team_game_number + month, family = "binomial", data = baseball)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7477  -1.1807   0.8089   1.0844   2.2635
```

```
Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                2.450e+00  6.793e-01   3.607  0.00031 ***
attendance                 2.324e-05  3.894e-06   5.967 2.42e-09 ***
time_of_game_minutes      -1.817e-02  1.901e-03  -9.559  < 2e-16 ***
day_night_indicatorN      -1.368e-01  1.109e-01  -1.233  0.21750
day_of_weekMon             3.072e-01  1.713e-01   1.794  0.07288 .
day_of_weekSat             1.156e-01  1.548e-01   0.747  0.45525
day_of_weekSun            -3.088e-02  1.790e-01  -0.173  0.86304
day_of_weekThu             2.902e-01  1.729e-01   1.678  0.09325 .
day_of_weekTue             3.629e-01  1.520e-01   2.388  0.01694 *
day_of_weekWed             2.707e-01  1.589e-01   1.704  0.08838 .
home_team_game_number     -4.540e-02  2.606e-02  -1.742  0.08146 .
visiting_team_game_number  4.279e-02  2.628e-02   1.628  0.10346
month                      8.742e-02  1.451e-01   0.602  0.54701
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3356.4  on 2428  degrees of freedom
Residual deviance: 3218.6  on 2416  degrees of freedom
  (1 observation deleted due to missingness)
AIC: 3244.6

Number of Fisher Scoring iterations: 4
```

```
m1_aug = augment(m1)
m1_aug <- m1_aug %>%
  mutate(prob = exp(.fitted)/(1 + exp(.fitted)),
         pred_home = ifelse(prob > 0.5, "home win", "not home win")) %>%
  select(.fitted, prob, pred_home, home_win)
table(m1_aug$pred_home, m1_aug$home_win)
```

```
               0   1
home win     612 939
not home win 521 357
```

```r
m1_aug$pred_home_num <- ifelse(m1_aug$pred_home == "home win", 1, 0)

accuracy <- sum(m1_aug$pred_home_num == m1_aug$home_win) / nrow(m1_aug)

accuracy
```

```
[1] 0.6010704
```