HW 7: NHL stats

Stats and sports class

Fall 2019

Preliminary notes for doing HW

1. All files should be knit and compiled using R Markdown. Knit early and often! I do not recommend waiting until the end of the HW to knit.
2. All questions should be answered completely, and, wherever applicable, code should be included.
3. If you work with a partner or group, please write the names of your teammates.
4. Copying and pasting of code is a violation of the Skidmore honor code

Homework questions

Part I: Readings

1. Read the summary model by the Evolving Wild twins:

https://rpubs.com/evolvingwild/395136/

Describe five unique hockey features that were implemented in their model. That is, look through their code, and highlight various ways that hockey-specific knowledge changed how they approached the problem.

2. Compare the three variable importance plots. Which variables were more important during even-strength play? Which were more important (relatively speaking) when a team was shorthanded or at uneven strength?

Part II: Implementation

We can access recent shot data here:

```
library(RCurl); library(tidyverse)
gitURL<- "https://raw.githubusercontent.com/statsbylopez/StatsSports/master/Data/pbp_data_hockey.rds"
pbp_data <- readRDS(gzcon(url(gitURL)))
names(pbp_data)
dim(pbp_data)
```

Question 1

Create a new variable for whether or not the shot occured during 5 v 5 play (that is, `home_skaters==5` and `away_skaters` == 5). Call this variable `is_5v5`.

Next, identify the goal rate (e.g, how often each shot was turned into a goal) within each cohort of `is_5v5`. That is, were shots more or less likely to go in during 5v5 play?

Question 2

Run the model below

```
library(broom)
fit_1 <- glm(event_type == "GOAL" ~ event_distance +
               event_angle + event_detail ,
```

```
              family = "binomial", data = pbp_data)
tidy(fit_1)
```

Interpret the coefficient on `event_detailWrist`

Question 3

Add `is_5v5` to your model in Question 2. Using AIC criterion, identify if this creates a preferable model.

Question 4

For `game_id == 2017020324`, identify each participating team's goals and expected goals. Did the outcome of this game match the relative shot inputs?

Bonus

Find the one game across the last two seasons where the different between the observed goal differential was as different from the expected goal differential