HW 7 SOLUTIONS: NHL stats

Stats and sports class

Fall 2020

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)))
```

Question 2

Interpret the coefficient on `event_detailWrist`

Solution: The odds of a successful shot are `exp(0.234) = 1.26` times higher on wrist shots, relative to backhand shots, using a model accounting for distance and angle. Note – the reference group is backhand shots and should be mentioned

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?

Solutions

```
pbp_data %>% filter(game_id == 2017020324) %>%
  group_by(event_team) %>%
  summarise(ave_xg = sum(shot_prob),
            act_goal = sum(event_type == "GOAL"))
```

Pittsburgh outscored Vancouver on expected goals 4.48 to 3.78, but lost 5 to 2

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

```
pbp_data %>%
  group_by(game_id) %>%
  summarise(xg_home = sum(shot_prob*(event_team == home_team)),
            xg_vis = sum(shot_prob*(event_team == away_team)),
            home_goal = sum((event_type == "GOAL")*(event_team == home_team)),
            vis_goal = sum((event_type == "GOAL")*(event_team == away_team))) %>%
  mutate(goal_diff = home_goal - vis_goal,
         xg_diff = xg_home - xg_vis,
         abs_diff = abs(goal_diff - xg_diff)) %>%
  arrange(-abs_diff) %>%
  head(5)
```

Solutions: game_id == 2018020141 was expected to be a 3.38 to 2.75 win for the home team, but was actually a 9 to 1 win for the away team. That has the largest difference in this data set. There are *lots* of ways to code this – any is sufficient

BONUS QUESTIONS WORTH +0.5 pts for those who got it right