adjusting for player effects

Jason Osborne and Rich Levine

2023-06-21

Batter Effects

We try to fit a logistic regression model for whether or not a home run is hit that includes effects for who the batter is. We will make it more complex in a bit. To avoid quasi-separation issues, we will condition away plate appearances by players who never hit any home runs. Or perhaps we will set them to a universal powerless batter.

```
load("allyrs.12vars.RData")
allyrs.12vars %>% select(BAT_ID,hr) %>% group_by(BAT_ID) %>%
  summarize(hrsum=sum(hr),pa=n()) -> batterHRs
batterHRs %>% filter(hrsum==0) -> batters0HRs
batters0HRs %>% dim; print(sum(batters0HRs$pa))
```

```
## [1] 1364    3
```

```
## [1] 42133
```

So, there are 1364 players who never hit a home run in an aggregate of 42133 plate appearances. Let us assign a new level of the `BAT_ID` factor to these players, maybe `TomOliver` after a Red Sox CF who went 2000 career plate appearances without ever hitting one over the fence.

```
batters0HRs %>% mutate(BAT_ID0="TomOliver") %>% select(BAT_ID,BAT_ID0,pa)-> batters0HRs
allyrs.12vars %>% left_join(batters0HRs) -> tmp
```

```
## Joining with `by = join_by(BAT_ID)`
```

```
tmp %>% mutate(BAT_ID=ifelse(BAT_ID0!="TomOliver",BAT_ID,"TomOliver")) %>% select(-"BAT_ID0") -> tmp
```

```
tmp %>% slice(1:3)
```

```
##         GAME_ID park season BAT_ID EVENT_CD   hr BAT_HAND_CD PIT_HAND_CD
## 1 ANA201204060  ANA   2012   <NA>        2 FALSE           L           R
## 2 ANA201204060  ANA   2012   <NA>        3 FALSE           R           R
## 3 ANA201204060  ANA   2012   <NA>        2 FALSE           L           R
##     PIT_ID pa
## 1 weavj003 NA
## 2 weavj003 NA
## 3 weavj003 NA
```

```
tmp %>% filter(BAT_ID=="TomOliver") %>% slice(1:3)
```

```
##         GAME_ID park season    BAT_ID EVENT_CD   hr BAT_HAND_CD PIT_HAND_CD
## 1 ANA201204220  ANA   2012 TomOliver        3 FALSE           R           R
## 2 ANA201204220  ANA   2012 TomOliver        2 FALSE           R           R
## 3 ANA201204300  ANA   2012 TomOliver        2 FALSE           L           L
##     PIT_ID pa
## 1 hared001 64
```

```
## 2 hared001 64
## 3 wilsc004 18
```