Adjusted Home Run Frequencies

Jason Osborne

Adjusted hr frequencies, 2022.

Let us see how the batter-pitcher combination (`bpcombo`) frequencies have varied over time

```
load("./allyrs.12vars.RData")
allyrs.12vars %>% select(BAT_HAND_CD,PIT_HAND_CD) %>% table %>% prop.table -> bpfrqs.era
# all four combo frqs
bpfrqs.era
```

```
##            PIT_HAND_CD
## BAT_HAND_CD          L          R
##           L 0.08054097 0.34107305
##           R 0.19685160 0.38153439
```

```
# batter hand
bpfrqs.era %>% rowSums()
```

```
##        L        R
## 0.421614 0.578386
```

```
# pitcher hand
bpfrqs.era %>% colSums()
```

```
##         L         R
## 0.2773926 0.7226074
```

```
# conditionally on batter hand, P(LHP|LHB), P(LHP|RHB)
bpfrqs.era[1,1]/sum(bpfrqs.era[1,])
```

```
## [1] 0.1910301
```

```
bpfrqs.era[2,1]/sum(bpfrqs.era[2,])
```

```
## [1] 0.3403464
```

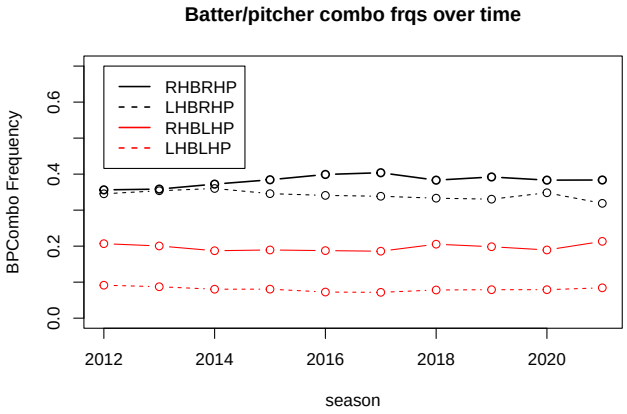
```
# conditionally on pitcher hand, P(LHB|LHP), P(LHB|RHP)
bpfrqs.era[1,1]/sum(bpfrqs.era[,1])
```

```
## [1] 0.2903501
```

```
bpfrqs.era[1,2]/sum(bpfrqs.era[,2])
```

```
## [1] 0.4720032
```

These four bpcombo frequencies have changed little over time, though the preference for LHB when facing RHP may have decreased slightly.

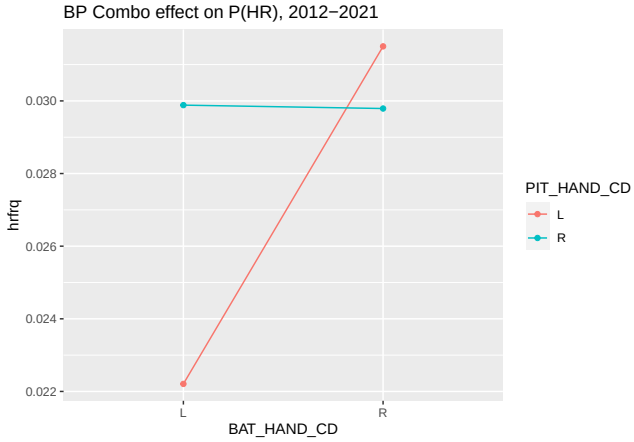


The effects of bpcombo on home run frequency can be investigated with an interaction plot, generated with data from 2012-2021:

```
allyrs.12vars %>% group_by(BAT_HAND_CD,PIT_HAND_CD) %>% summarize(hrfrq=mean(hr)) -> hrfrq.bp.era
hrfrq.bp.era
```

```
## # A tibble: 4 x 3
## # Groups:   BAT_HAND_CD [2]
##   BAT_HAND_CD PIT_HAND_CD  hrfrq
##   <chr>       <chr>        <dbl>
## 1 L           L           0.0222
## 2 L           R           0.0299
## 3 R           L           0.0315
## 4 R           R           0.0298
```

```
ggplot(hrfrq.bp.era,aes(y=hrfrq,x=BAT_HAND_CD,color=PIT_HAND_CD)) +
  geom_line(aes(group=PIT_HAND_CD)) + geom_point() +
  ggtitle("BP Combo effect on P(HR), 2012-2021")
```



Looking over this 10 year period, it can be seen that home runs are least likely when a LHB is facing a LHP. Remarkably, the effect of the batter hand only appears to matter when facing lefties. Wow!

For a given park, the frequencies of the four combinations can vary dramatically from one season to the next, depending upon the personnel of the home team and with the unbalanced schedules of years past, upon the personnel of other teams in the division. In light of bpcombo effects, home run frequencies for each park can be adjusted to league-wide bpcombo frequencies simply by reweighting the four conditional home run rates

to the these frequencies.

```
allyrs.12vars %>% group_by(park,BAT_HAND_CD,PIT_HAND_CD) %>%
  summarize(hrfrq=mean(hr)) %>% pivot_wider(values_from=hrfrq,
                                names_from=c("BAT_HAND_CD","PIT_HAND_CD")) ->
  hrsummary.wide
hrsummary.wide %>% mutate(adjhr=bpfrqs.era[1,1]*L_L + bpfrqs.era[1,2]*L_R +
                                bpfrqs.era[2,1]*R_L + bpfrqs.era[2,2]*R_R) ->
  hrsummary.wide
# unadjusted
allyrs.12vars %>% group_by(park) %>% summarize(obshr=mean(hr)) -> hrfrqs.bypark
hrsummary.wide %>% inner_join(hrfrqs.bypark) -> hrsummary.wide
```

These adjusted frequencies can be plotted against park, along with the unadjusted frequencies. Further investigation of changes over time is warranted though, as a glance at 10-year averages still shows considerable variability in bpcombo frequencies across parks. It must be kept in mind that many players reside with the same team for long periods of time, so these 10 years are not at all independent. However, we average anyway ...

A technique worth mentioning in the construction of this plot is to achieve an ordering of parks on the horizontal axis according to either the observed or adjusted home run rate by so ordering the levels of park as a factor.

```
hrsummary.wide$park <- factor(hrsummary.wide$park,
                              levels=hrsummary.wide$park[order(hrsummary.wide$adjhr)])
hrsummary.wide %>% pivot_longer(cols=c("adjhr","obshr"),values_to="phr") ->
  hrsummary.tall
```

Now `ggplot` can be used ...

```
ggplot(hrsummary.tall) + geom_point(aes(y=phr,x=park,color=name)) +
  theme(axis.text.x=element_text(angle=50)) +
  ggtitle("HR Freqs, adj and observed \n 2012-2021")
```