Lesson_12_Boardsheet

Kevin Cummiskey

2/7/2020

Review

Last class, we discussed the Run Expectancy Matrix and how to obtain it from the Retrosheet play-by-play data. Here is the Run Expectancy Matrix from last class.

```
library(tidyverse)
library(knitr)
library(Lahman)

RUNS <- read_csv(file = "./data/expectedRUNs2016.csv")

RUNS_out = matrix(round(RUNS$Mean,2), 8,3)
colnames(RUNS_out) = c("0 outs", "1 out", "2 outs")
rownames(RUNS_out) = c("000","001","010","011",
                       "100","101","110", "111")
kable(RUNS_out)
```

	0 outs	1 out	2 outs
000	0.50	0.27	0.11
001	1.35	0.94	0.37
010	1.13	0.67	0.31
011	1.93	1.36	0.55
100	0.86	0.51	0.22
101	1.72	1.20	0.48
110	1.44	0.92	0.41
111	2.11	1.54	0.70

Let's say there is a runner on 2nd base. Stealing 3rd base is the most valuable when there are how many outs?

Measuring the Success of a Batting Play

We can estimate the value of a plate appearance as the difference in run expectancies between the new and old states plus the number of runs scored on the play.

$$\text{RUN VALUE} = \text{RUNS}_{\text{new state}} - \text{RUNS}_{\text{old state}} + \text{RUNS}_{\text{scored on play}}$$

Calculate the RUN VALUE of the following plays:

(1) There are runnners on first and second with no outs. The batter successfully sacrifice bunts, resulting in runners on second and third with 1 out.

(2) There is a runner on second base with no outs. The batter hits a single that scores the runner from second. The batter stops at first base.

(2) There is a runner on second base with two outs. The batter hits a single that scores the runner from second. The batter stops at first base.

(2) What is the most valuable play in baseball? Explain.

Here's how to calculate the run value suing the code from last class (pg 116).

```
# load data2016 (as of page 115 in text)
data2016 <- read_csv(file = "./data/data2016.csv")

# add run value of the play (page 116)
data2016 %>%
  left_join(select(RUNS, -Outs), by = "STATE") %>%
  rename(Runs.State = Mean) %>%
  left_join(select(RUNS, -Outs),
            by = c("NEW.STATE" = "STATE")) %>%
  rename(Runs.New.State = Mean) %>%
  replace_na(list(Runs.New.State = 0)) %>%
  mutate(run_value = Runs.New.State - Runs.State +
           RUNS.SCORED) -> data2016
```

Let's look at Jose Altuve (page 117).

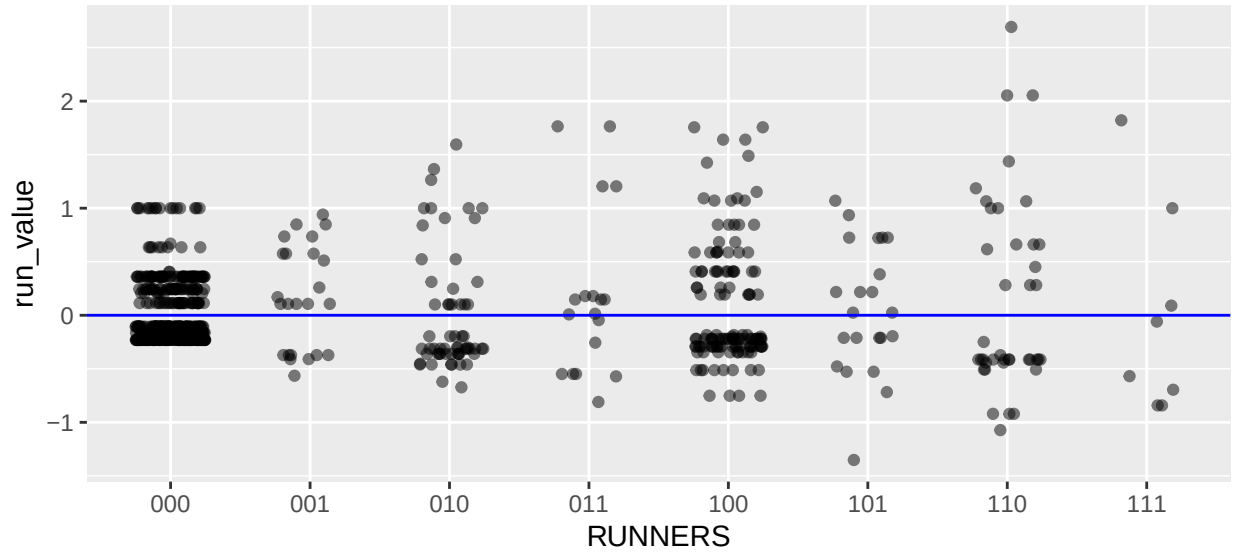
```
#get Altuve's playerID
Master %>%
  filter(nameFirst == "Jose", nameLast == "Altuve") %>%
  pull(retroID) -> altuve.id

data2016 %>%
  filter(BAT_ID == altuve.id,
         BAT_EVENT_FL == TRUE) -> altuve

altuve %>%
  select(STATE, NEW.STATE, run_value) %>%
  slice(1:3)
```

```
## # A tibble: 3 x 3
##   STATE NEW.STATE run_value
##   <chr> <chr>         <dbl>
## 1 000 1 000 2        -0.162
## 2 000 1 100 1         0.244
## 3 000 1 000 2        -0.162
```

Here's a plot of Altuve's run value by runner state.

```
altuve %>%
  ggplot(aes(x = BASES, y = run_value)) +
  geom_jitter(width = 0.25, alpha = 0.5) +
```

```
  geom_hline(yintercept = 0, color = "blue") +
  xlab("RUNNERS")
```



Next, let's look at total runs by runners.

```
altuve %>%
  group_by(BASES) %>%
  summarize(RUNS = sum(run_value),
            PA = n()) -> Runs_Altuve
Runs_Altuve
```

```
## # A tibble: 8 x 3
##   BASES    RUNS    PA
##   <chr>   <dbl> <int>
## 1 000   10.1      417
## 2 001    4.06      24
## 3 010    0.0695    60
## 4 011    3.43      18
## 5 100   10.2      128
## 6 101    1.34      22
## 7 110    5.62      40
## 8 111   -0.0968     8
```

Lastly, we can calculate a statistic called RE24 which sums RUNS over the different base runner combinations.

```
Runs_Altuve %>% summarize(RE24 = sum(RUNS))
```

```
## # A tibble: 1 x 1
##    RE24
##   <dbl>
## 1  34.7
```

How do we interpret RE24?

In baseball, there are context-neutral and context-dependent statistics. Which is RE24?