Lesson_13_Boardsheet

Kevin Cummiskey

2/12/2020

Review

Last class, we discussed the run value of a play. Write the equation for the run value of a play.

Describe how you would calculate the RE24 for David Ortiz in 2016. Also, interpret this value in a way that's understandable to a general audience.

Is it appropriate to compare David Ortiz's and Jose Altuve's RE24? Explain.

Value of a home run

Let's look at the value of a home run from a runs perspective. First, here is the distribution of home runs by the number of runs scored.

```
library(tidyverse)
library(Lahman)

#recall data2016 contains a line for every play
data2016 <- read_csv("data/data2016.csv")

#get only home runs
data2016 %>%
  filter(EVENT_CD == 23) -> home_runs

#number and percent of homeruns by runs scored
home_runs %>%
  group_by(RUNS.SCORED) %>%
  summarise(n = n()) %>%
  mutate(perc = round(n/sum(n),2))
```

```
## # A tibble: 4 x 3
##   RUNS.SCORED     n  perc
##         <dbl> <int> <dbl>
## 1           1  3332  0.59
## 2           2  1534  0.27
## 3           3   634  0.11
## 4           4   110  0.02
```

```
#average number of runs scored
home_runs %>%
  summarise(mean(RUNS.SCORED))
```

```
## # A tibble: 1 x 1
##   `mean(RUNS.SCORED)`
##                 <dbl>
## 1                1.56
```

Explain why we might want to use the run value approach of Chapter 5 instead of the average runs scored approach above.

```
#calculate the average run value of a home run
mean_hr <- home_runs %>%
  summarise(mean_run_value = mean(run_value))
mean_hr
```

```
## # A tibble: 1 x 1
##   mean_run_value
##            <dbl>
## 1           1.38
```

Why do the two approaches result in different answers?

```
home_runs %>%
  ggplot(aes(run_value)) +
  geom_histogram() +
  geom_vline(data = mean_hr, aes(xintercept = mean_run_value),
             color = "blue", size = 1.5) +
  annotate("text", 1.7,2000, label = "Mean Run\nValue",
           color = "blue")
```