

Lesson_26_Boardsheet - Equivalence Coefficient

Kevin Cummiskey

3/30/2020

Don Mattingly (“Donnie Baseball”) - Should he be in the Hall of Fame?



“Mattingly was arguably the best player in Major League Baseball for a six-year span from 1984-89. He averaged 27 homers and 114 RBIs, while putting up a .327/.372/.530 batting line in those six campaigns. The .530 slugging percentage led all qualified hitters, while Mattingly also paced the Majors in extra-base hits (428) and RBIs (684) during that six-year span.”

“Though he hit a respectable 222 homers, Mattingly wasn’t just a power-hitting first baseman. He finished his career with a .307 average, making him one of just eight first basemen in Major League history to hit at least .305 with 200 homers. That list includes five Hall of Famers in Lou Gehrig, Jimmie Foxx, Johnny Mize, Hank Greenberg and Jim Bottomley.”

“Though Mattingly was one of the game’s elite talents from 1984-89, he was an entirely different player starting with the ‘90 campaign in which he was limited to just 102 games due to a congenital disk deformity in his back. Mattingly averaged only 10 homers and 64 RBIs over his final six seasons, while hitting .286/.345/.405. He topped out at 17 homers and 86 RBIs during that stretch. Though his superb defense continued throughout his career, Mattingly ended his career by hitting .288 with seven homers and 49 RBIs over 128 games in ‘95.”

Excerpts from *Here’s the HOF case for and against Mattingly* by Paul Casella <https://www.mlb.com/news/don-mattingly-hall-of-fame-case>)

Career Statistics

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

```
vars <- c("G", "AB", "R", "H", "X2B", "X3B",
          "HR", "RBI", "BB", "SO", "SB", "SH", "HBP", "SF")
Batting %>%
  filter(playerID == "mattido01") %>%
  summarise_at(vars, sum) %>%
  mutate(AVG = round(H/AB,3)) %>%
  select(G, AB, H, HR, AVG) %>%
  kable(caption = "Mattingly Career Statistics")
```

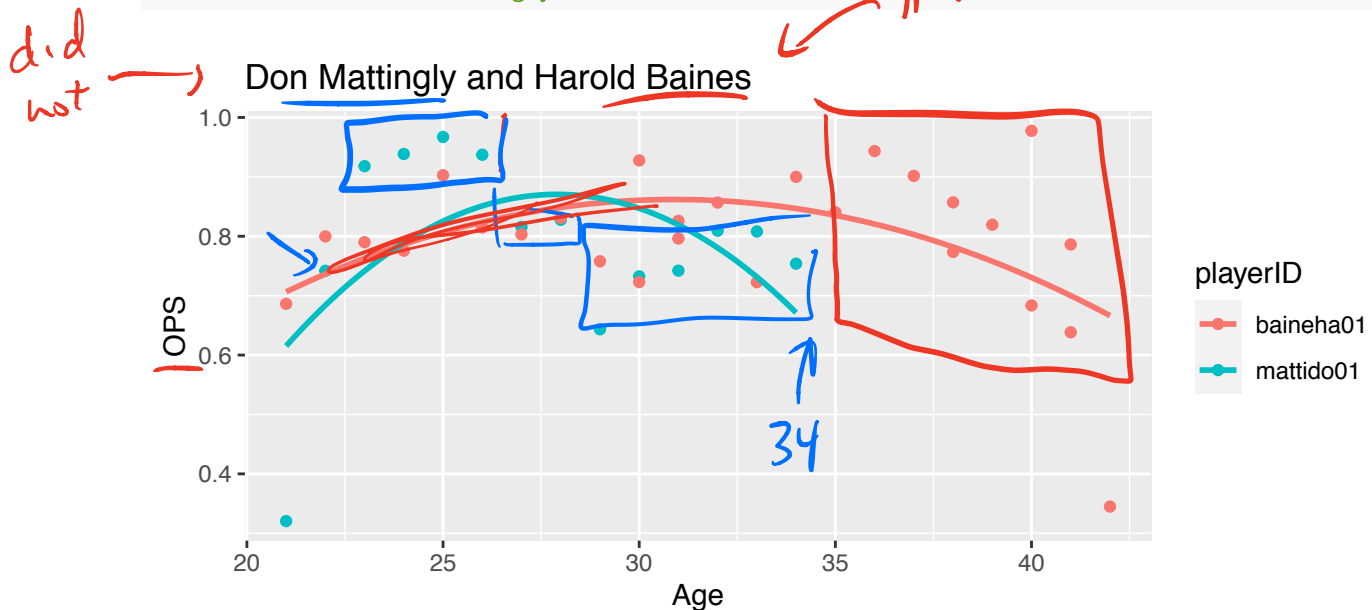
Table 1: Mattingly Career Statistics

G	AB	H	HR	AVG
1785	7003	2153	222	0.307

Creates
 H > 3000
 HR > 500

A picture's worth a thousand words.

```
source("Chapter8_functions.R")
stats <- c("mattido01", "baineha01") %>% map_df(get_stats)
stats %>%
  ggplot(aes(x = Age, y = OPS, color = playerID)) +
  geom_point() +
  geom_smooth(method = "lm",
              formula = y ~ x + I(x^2),
              se = FALSE) +
  labs(title = "Don Mattingly and Harold Baines")
```



How does he compare to other Hall of Famers?

```
library(tidyverse)
library(Lahman)
library(glue)
library(plotly)
library(ggrepel)

vars <- c("G", "AB", "R", "H", "X2B", "X3B",
          "HR", "RBI", "BB", "SO", "SB",
          "SH", "HBP", "SF")

#Career totals for players with 5000 at bats
Batting %>%
  group_by(playerID) %>%
  summarize_at(vars, sum, na.rm = TRUE) %>%
  filter(AB > 5000) %>%
  mutate(AVG = round(H/AB, 3),
         HR.rate = HR/AB,
         PA = AB + BB + SH + HBP + SF) -> career.totals

# Must have retired at least five years ago
Master %>%
  filter(finalGame < 2014,
         finalGame > 1920) %>%
  pull(playerID) -> eligible
career.totals %>%
  filter(playerID %in% eligible) -> career.totals

#determine whether they are in the hall of fame
career.totals %>%
  left_join(HallOfFame %>%
    filter(inducted == "Y",
           category == "Player") %>%
    select(playerID, yearID, inducted),
           by = "playerID") %>%
  replace_na(list(inducted = "N")) %>%
  rename(halloffame = inducted) -> career.totals

#add full name
career.totals %>%
  left_join(Master %>% select(nameLast, nameFirst, playerID)) %>%
  mutate(name = paste(nameFirst, nameLast, sep = " ")) %>%
  select(-nameLast, -nameFirst) -> career.totals

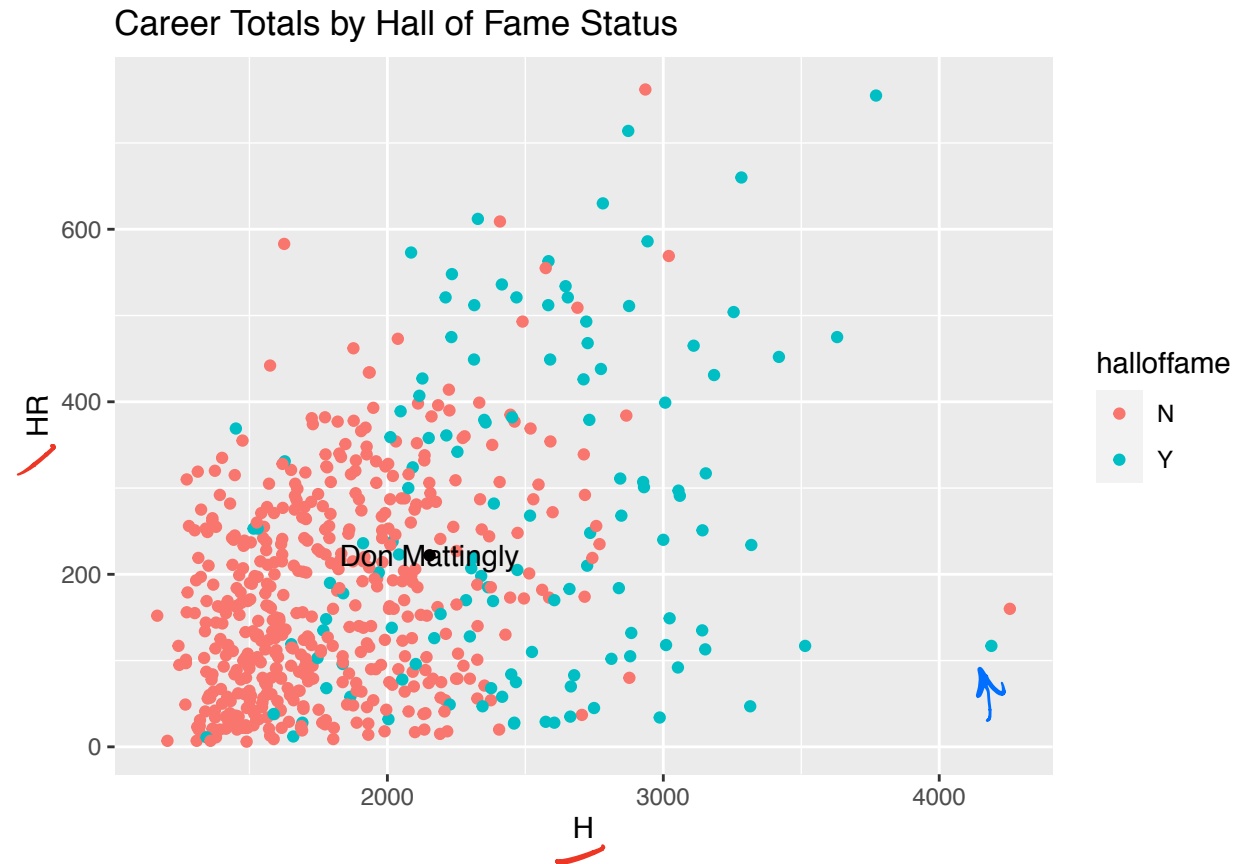
# Career Totals

p2 <- career.totals %>%
  ggplot(aes(label = name, x = H, y = HR, color = halloffame)) +
  geom_point() +
  geom_text(data = filter(career.totals, playerID == "mattido01"),
           aes(label = name), color = "black") +
```

HallOfFame

```
geom_point(data = filter(career.totals, playerID == "mattido01"),
  aes(x = H, y = HR), color = "black") +
labs(title = "Career Totals by Hall of Fame Status")

#ggplotly(p2)
p2
```



A logistic regression model predicting hall of fame selection.

"Predict Hall of Fame induction"

```
# Here's a model based on aggregate statistics for the probability of Hall of Fame
library(mgcv)
model.hof <- glm(halloffame == "Y" ~ H + X2B + X3B + HR + RBI + BB,
  family = "binomial",
  data = career.totals)

career.totals %>%
  mutate(hof.pred = predict(model.hof,
    type = "response",
    newdata = .)) -> career.totals

career.totals %>%
  arrange(desc(hof.pred)) %>%
  select(name, hof.pred, halloffame) %>%
  head(10) %>%
  kable(caption = "Top Ten Highest Predicted Probability of Hall of Fame")
```

Table 2: Top Ten Highest Predicted Probability of Hall of Fame ✓

	name	hof.pred	halloffame
→	Ty Cobb	0.9999104	Y
	Eddie Collins	0.9948654	Y
✓	Babe Ruth	0.9944104	Y
✓	Stan Musial	0.9942120	Y
✓	Hank Aaron	0.9935800	Y
	Tris Speaker	0.9880107	Y
✓	Lou Gehrig	0.9875152	Y
✓	Willie Mays	0.9857596	Y
	Pete Rose	0.9785629	N
	Jimmie Foxx	0.9720099	Y

```
career.totals %>%
  arrange(desc(hof.pred)) %>%
  filter(hof.pred > 0.061, hof.pred < 0.064) %>%
  select(name, hof.pred, halloffame) %>%
  kable(caption = "Players with similar predictions as Don Mattingly")
```

Table 3: Players with similar predictions as Don Mattingly

	name	hof.pred	halloffame
	Milt Stock	0.0638950	N
	Howie Shanks	0.0636585	N
✓	Kent Hrbek	0.0632855	N
	Rick Monday	0.0631722	N
	Don Mattingly	0.0621394	N
	Ron Gant	0.0620645	N
✓	Jim Edmonds	0.0616916	N
	Harlond Clift	0.0615806	N
	Roy White	0.0612990	N

6%

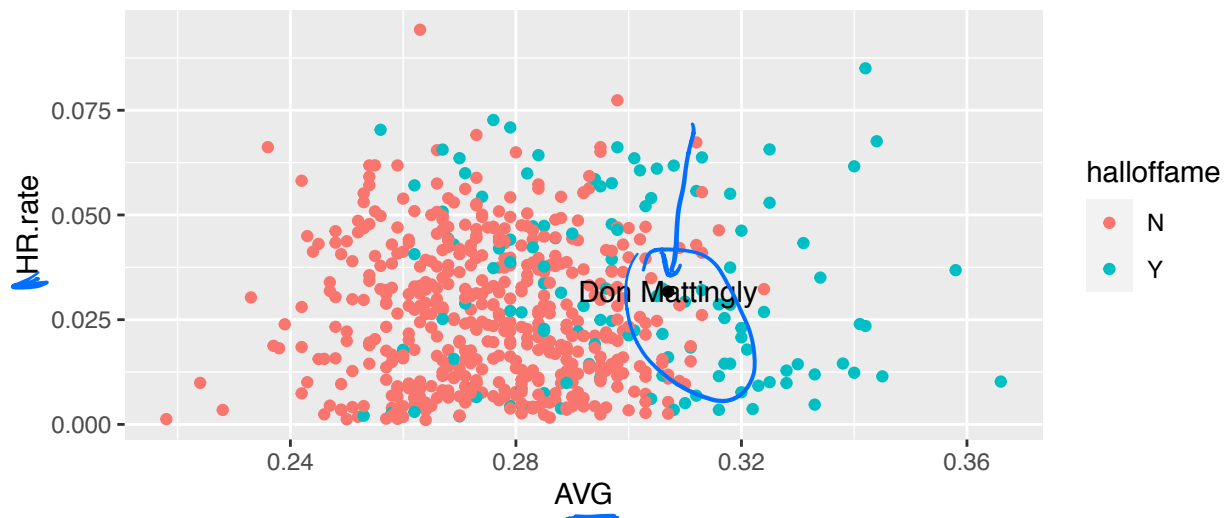
Let's compare Don Mattingly and Harold Baines.

```
# Relative Statistics

p1 <- career.totals %>%
  ggplot(aes(label = name, x = AVG, y = HR.rate, color = halloffame)) +
  geom_point() +
  geom_text(data = filter(career.totals, playerID == "mattido01"),
    aes(label = name), color = "black") +
  geom_point(data = filter(career.totals, playerID == "mattido01"),
    aes(x = AVG, y = HR.rate), color = "black") +
  labs(title = "Career Totals by Hall of Fame Status")

#ggplotly(p1)
p1
```

Career Totals by Hall of Fame Status



```
career.totals %>%
  filter(playerID %in% c("mattido01", "baine01")) %>%
  select(name, G, PA, AB, H, HR, AVG) %>%
  kable(caption = "Career Statistics")
```

Table 4: Career Statistics

name	G	PA	AB	H	HR	AVG
Harold Baines	2830	11092	9908	2866	384	0.289
Don Mattingly	1785	7721	7003	2153	222	0.307

HOF
→
Equivalence Coefficient

"What if"

← Baines and Mattingly had + the number of PA.

Now, let's assume Mattingly and Baines both had 12000 plate appearances. Following the steps on page 47 of *Understanding Sabermetrics*, let's compare cumulative statistics for various hitters.

- Determine the projected number of new at bats. 12000
- Solve for x in $\frac{AB}{AB+BB} = \frac{AB+x}{newPA}$. This is the number of additional at bats to award each hitter.

```
newPA = 12000
career.totals %>%
  mutate(x = (AB/(AB + BB))*newPA - AB) -> career.totals

# x is the number of additional at bats
career.totals %>%
  filter(playerID %in% c("mattido01", "baine01")) %>%
  select(name, AB, x) %>%
  kable(caption = "Number of new at bats.")
```

Table 5: Number of new at bats.

name	AB	x
Harold Baines	9908	930.2862
Don Mattingly	7003	4067.4782

number of at bats
to add to player's
total

- Calculate the equivalence coefficient, $(1 + \frac{x}{AB} * k)$, where k is the kicker. Let's investigate $k = 1, 0.95, 0.90$.

$k = \text{kicker}$

#function to compute equivalence coefficient

```
eqcoef <- function(k, d){
  d %>%
    mutate(k = k,
           EC = 1 + x/AB * k,
           H = round(H*EC,0),
           X2B = round(X2B * EC,0),
           X3B = round(X3B * EC,0),
           HR = round(HR * EC,0),
           RBI = round(RBI * EC,0),
           BB = round(BB * EC,0)
    ) -> d
  return(d %>% select(playerID, name, k, EC, H, X2B, X3B, HR, RBI, BB))
}
```

$k = c(1, 0.95, 0.9)$

$k \%>\%$
map_df(eqcoef, d = career.totals %>% filter(playerID == "mattido01")) -> mattie.ec

$k \%>\%$
map_df(eqcoef, d = career.totals %>% filter(playerID == "baine01")) -> baine.ec

rbind(mattie.ec, baine.ec) %>%
kable()

If Mattingly had not been

injured,
there would
have been
a strong
case for the HOF

playerID	name	k	EC	H	X2B	X3B	HR	RBI	BB
mattido01	Don Mattingly	1.00	1.580819	3404	699	32	351	1737	930
mattido01	Don Mattingly	0.95	1.551778	3341	686	31	344	1705	912
mattido01	Don Mattingly	0.90	1.522738	3278	673	30	338	1673	895
baine01	Harold Baines	1.00	1.093892	3135	534	54	420	1781	1162
baine01	Harold Baines	0.95	1.089198	3122	532	53	418	1773	1157
baine01	Harold Baines	0.90	1.084503	3108	529	53	416	1766	1152

Using his projected career results, let's see how likely he would have been to make the Hall of Fame.

```
mattie.ec %>%
  mutate(hof.pred = predict(model.hof,
                             type = "response",
                             newdata = .)) -> mattie.ec
```

61.

```
matty.ec %>%
  select(name, k, hof.pred) %>%
  kable()
```

name	k	hof.pred
Don Mattingly	1.00	0.7174097
Don Mattingly	0.95	0.6771133
Don Mattingly	0.90	0.6343150

```
#players similar at 'hypothetical mattingly'
career.totals %>%
  filter(hof.pred > 0.60, hof.pred < 0.65) %>%
  select(name, hof.pred, halloffame) %>%
  kable()
```

name	hof.pred	halloffame
Willie Davis	0.6158938	N
Tony Gwynn	0.6281987	Y
Chipper Jones	0.6068336	Y
Stuffy McInnis	0.6380384	N
Vada Pinson	0.6087094	N
Tim Raines	0.6247221	Y
Jim Rice	0.6416950	Y
Brooks Robinson	0.6438496	Y
Mike Schmidt	0.6048885	Y
Rusty Staub	0.6416452	N
Bobby Veach	0.6489021	N

Review

- Equivalence Coefficient - used to compare aggregate statistics of players with different number of PA.

1. Question: did the called strike rate decrease?

2. Data: restrict players
p = probability of called strike

$$\log\left(\frac{p}{1-p}\right) =$$

taken

- leave swinging strikes in

- all strikes

"swing"

Ball - B

Called Strike - S }
Swung Strike - X }

swung
and called