

Outline
o

Intro
ooo

Data acquisition
ooo

Modeling clutch
oooooooo

Factors affecting clutch
ooooo

Group comparison
ooo

Conclusions
ooo

Clutch batting in MLB

Hao-Wei Chu

July 25,2021

What is clutch and how do we find them

Data acquisition

Using statistical method to evaluate clutch batting ability

Searching for factors affecting clutch batting ability

Comparing “clutch ability” among groups

Conclusions

What is clutch batting in baseball?

- Clutch is the situations when a player at bat have a higher chance to change the game result.
- Our definition (from retrosheet.org)

Men on bases

RISP

Late & close

Bases loaded

- Does clutch hitting ability exists?
 - Virtually no evidence on clutch hitting [Cramer1977]
 - Exists, but the difference should be negligible [Silver2006]

Potential applications

- For team managers:
To provide measures which affects the performance of a player in critical situation, and help making decision which leads toward winning close games
- For sport betting business:
To enhance the accuracy in predicting games, which could help winning more money from the customers

The goal of the project

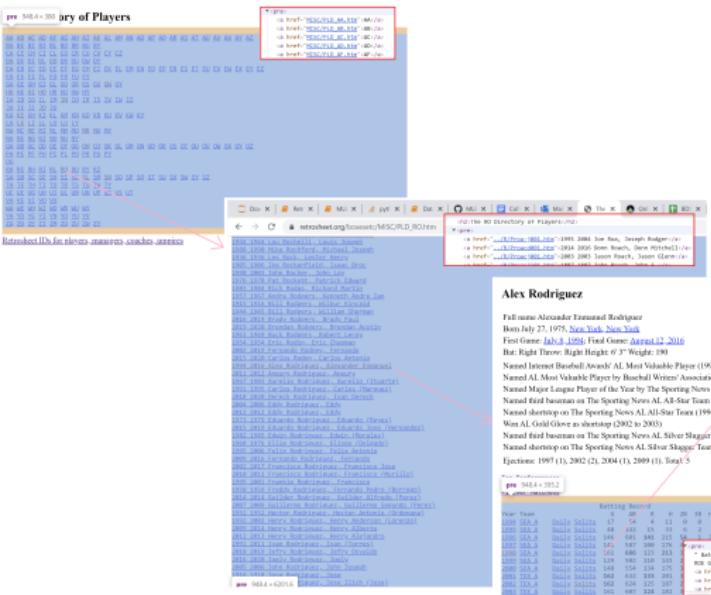
- Scrape batting data from <http://retrosheet.org>, and collect career and season data of all the MLB players in the history.
- Use the traditional data collected from <http://retrosheet.org>, describe measures for clutch batting ability.
- Search for (non)factors which affects clutch batting ability.
- Find evidences which supports or refutes the hypothesis that clutch batting ability does not exist.

<http://retrosheet.org>

The information used here was obtained free of charge from and is copyrighted by Retrosheet. Interested parties may contact Retrosheet at "<http://retrosheet.org>".

Scraping data from retrosheet

We need to walk through four layers before reaching the desired data.



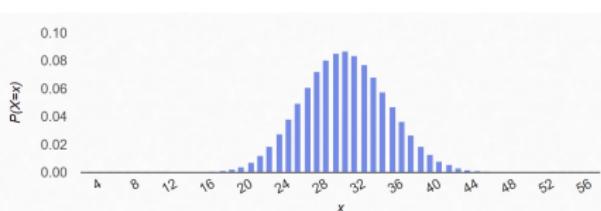
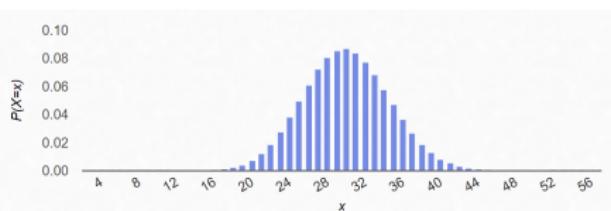
- Most of the beef (links, data) are embedded in the HTML pre tag with nice formatting.
Currently, there are
 - 19917 player overview pages to visit
 - 16009 career split data pages to scrape
 - 84231 season split data pages to pick up
- Some loose ends:
- Will get occasional 404 errors
- There are some discrepancies, marked with an i, which needs some care.
- Some statistics are not available to old players

How do we capture “clutch”?

- If clutch ability does not exist, how do we formulate this?
- The hypothetical situation: Suppose we get a group of people, throwing an unfair coin $N = 100$ times, with probability $p = 0.3$ of head each time...

- Theoretical

- Existence of “coin controllers” ...

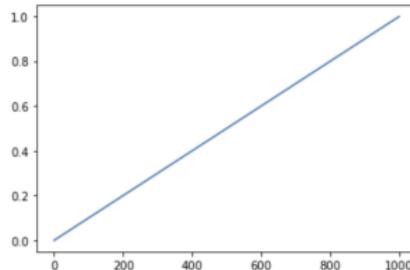
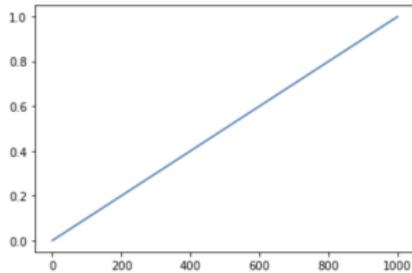


Ideas for normalization

- The reality is, not every player have the same coin...
- To balance the effect, we consider the p -value for each distribution instead.

- Theoretical

- Existence of “coin controllers” ...



Other issues

- Retrosheet provides batting data for 4 clutch situations:
 - When there are men on base
 - When there are runners in scoring position (RISP)
 - “Late and close”
 - When bases are loaded
- Looking at three classic batting numbers: AVG, OBP, and SLG
- In each situation, conduct a linear regression to balance the statistics difference in and not in that situation
 - e.g. the $\text{AVG}_{\text{non RISP}}$ is .2626, and AVG_{RISP} is .2655, so we apply a balancing factor $\theta = 1.0111$.

Example

Name	AVG	AVG _{RISP}	p_{RISP}	p_{MenOn}	p_{CnL}	p_{Loaded}
Ivan Rodriguez	.293	.290	.139	.271	.266	.520
Hank Aaron	.296	.326	.994	.767	.997	.485

The “non-existence of clutch ability” hypothesis

- Taking the RISP situation as possible, after the regression, the null hypothesis we will make on the metric AVG is:

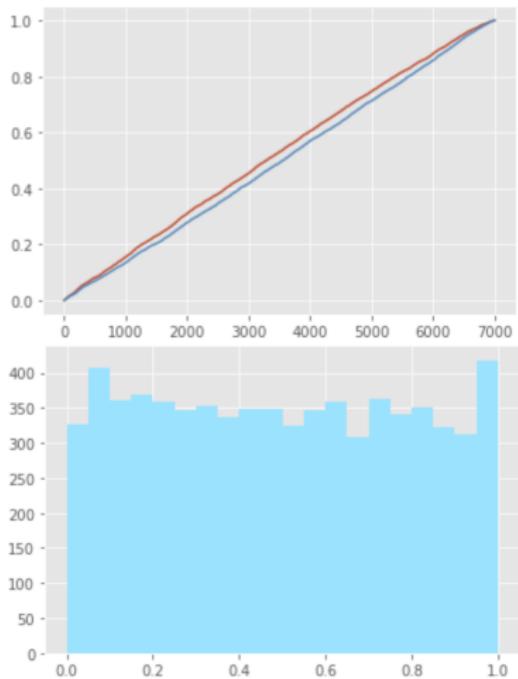
$$H_0 : \text{AVG}_{\text{non-RISP}} = \text{AVG}_{\text{RISP}},$$

for all qualified players in the database.

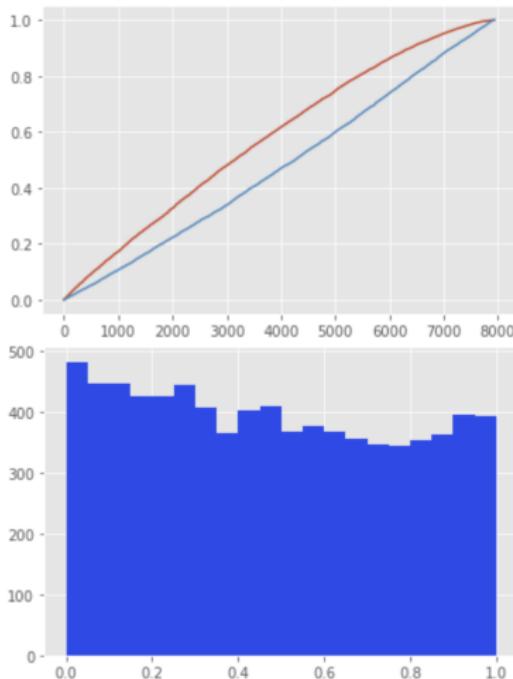
- Under the null hypothesis, the normalized p -value should be a uniform distribution in $(0, 1)$.
- We apply the Kolmogorov–Smirnov test to check whether the experimental distribution is uniform.

Test results:

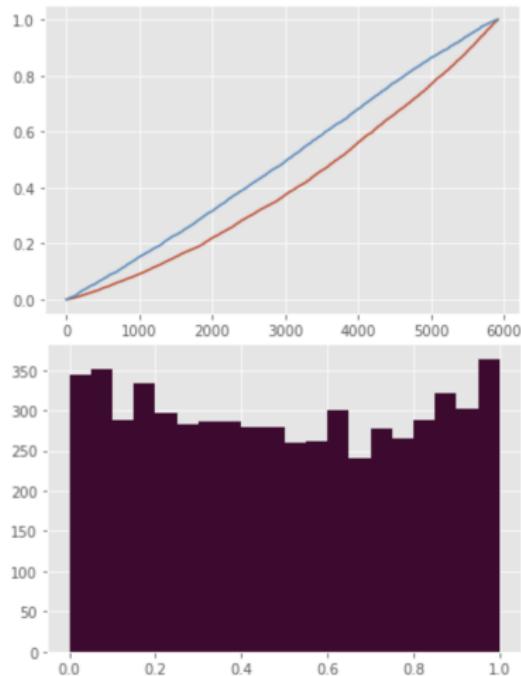
- RISP ($p_{KS} = 0.1356$)



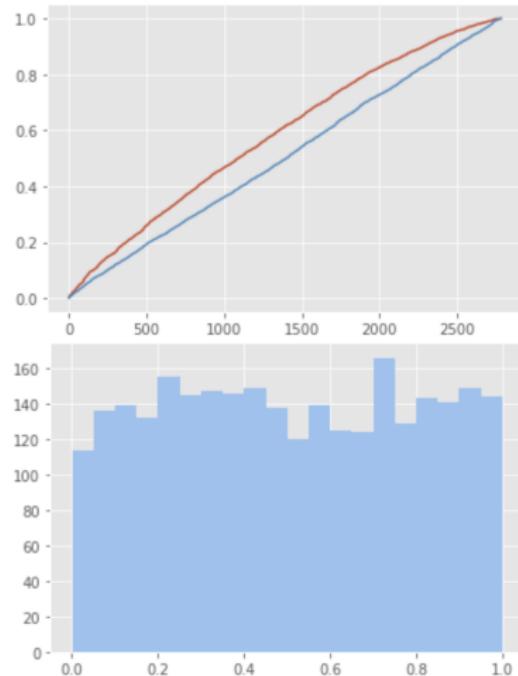
- Men on bases ($6.97e-12$)



- Close and late
($p_{KS} = 0.0012$)



- Bases loaded
($p_{KS} = 0.5413$)

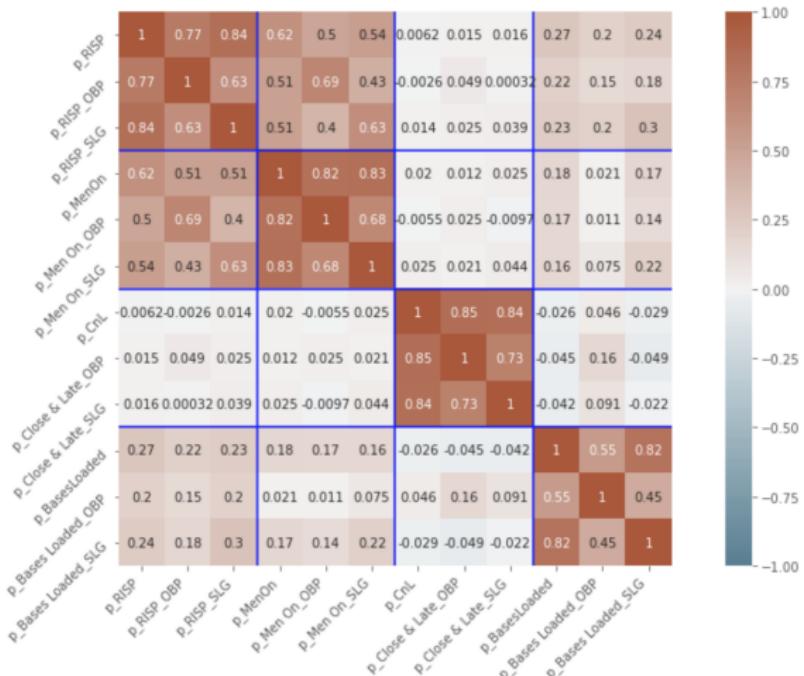


A discussion

- The null hypothesis is not rejected in some situations and rejected in other.
- The null hypothesis tends to pass when the correction factor is within $1 \pm 2\%$.
- We can not tell if batting in clutch situation is not pure randomness, or we need better regression models.



The correlations between the p -values

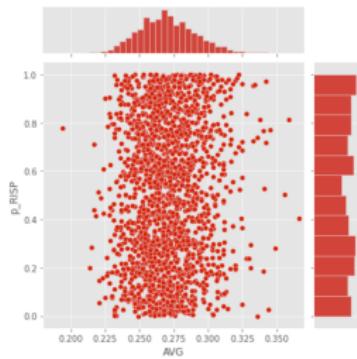


- Note that the “Close and Late” metrics have very low correlation with the other 3, which more or less overlaps.

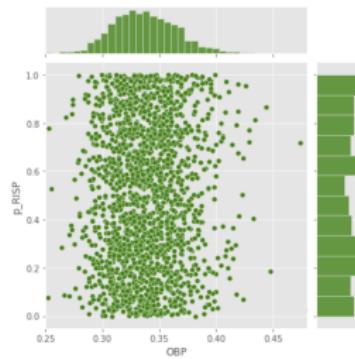
The correlations to AVG, OBP, and SLG

Seemed to have no correlation!

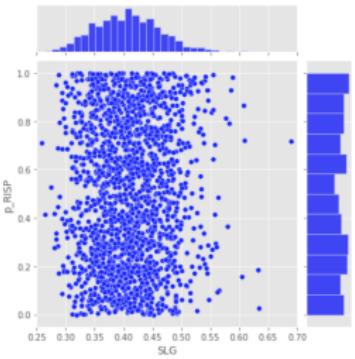
AVG



OBP



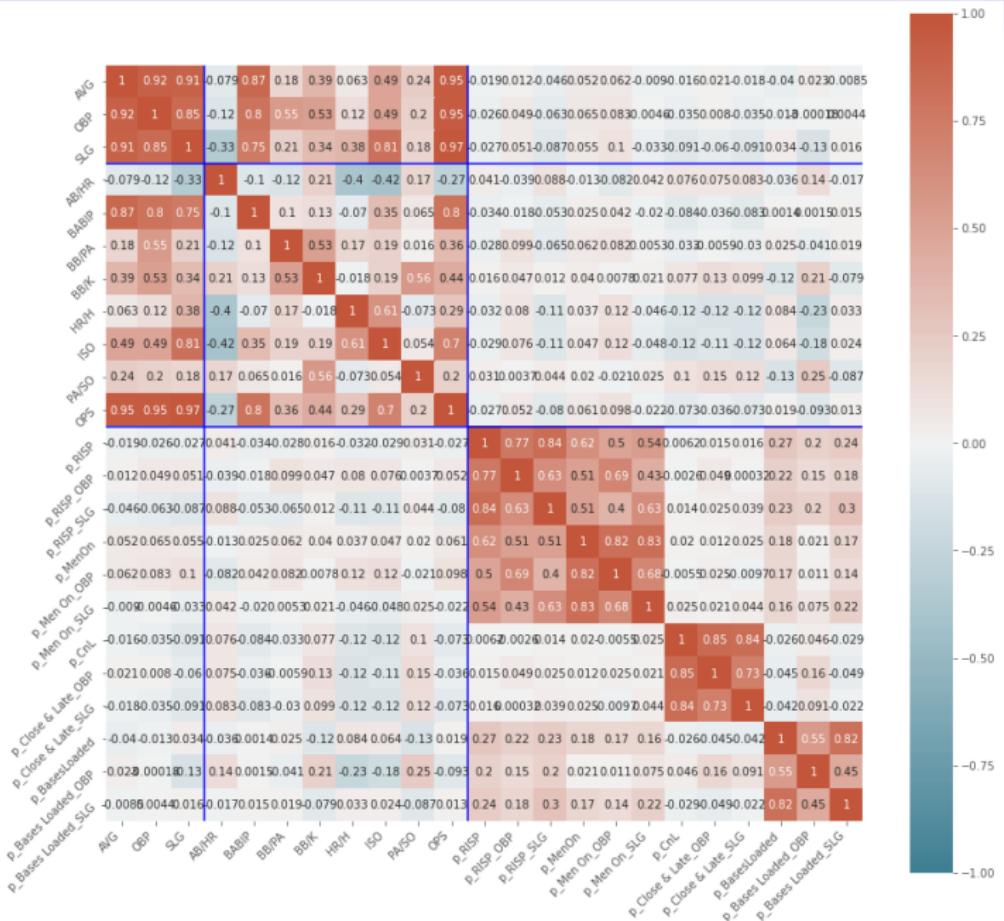
SLG



Other possible factors

We also took a look at other factors, such as

- AB/HR: At bats per home run
- BABIP: Batting average on balls in play
- BB/PA: The frequency to get a walk
- BB/K: Number of walks per strike out
- HR/H: Home runs per hit
- ISO: Isolated power: a hitter's ability to hit for extra bases
- PA/SO: Plate appearances per strikeout
- OPS: On-base average plus slugging average
- ... much more to list in baseball statistics



What can we tell?

All the correlations between basis statistics and the modified p -values are statistically insignificant (< 0.2), so predicting might be infeasible...

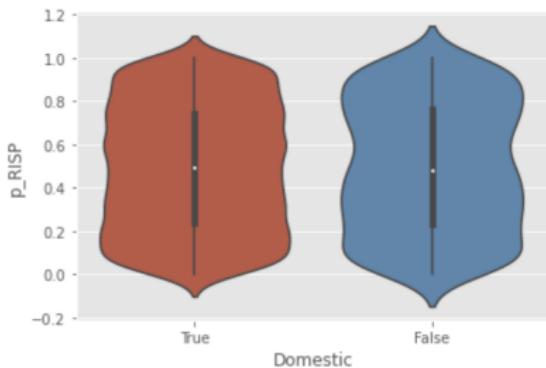
But if we really need to say something (by looking at correlations on items around 0.1 to 0.2):

- Sluggers are worse in clutch situations
- Home plate discipline: Players getting struck out less and picks more walks are better in clutch situations.
(this is also mentioned in Nate Silver's article)

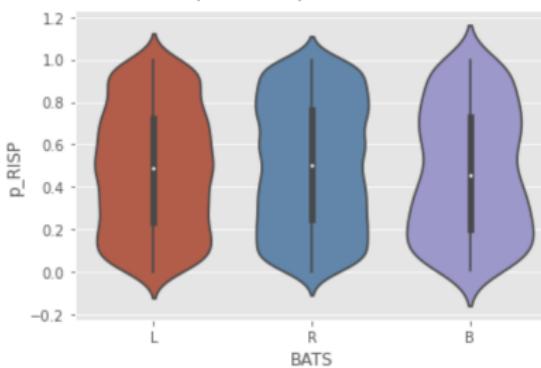
Categorizing by nationality and batting type

Nothing significant...

US v.s. international

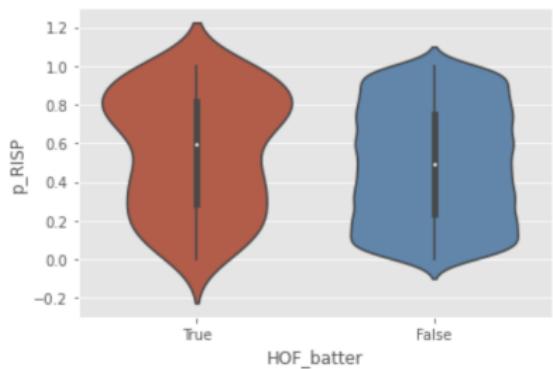


Left/right/switch

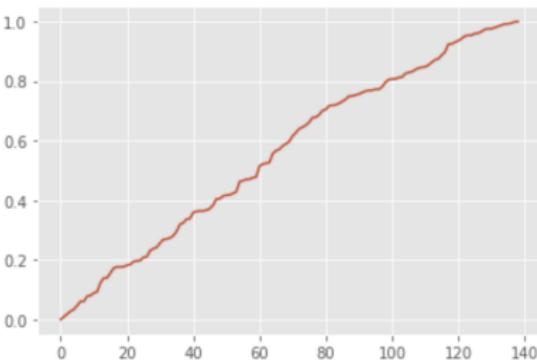


The case of Hall of Famers

Hall of Famers



The CDF of HoFers ($N = 140$)

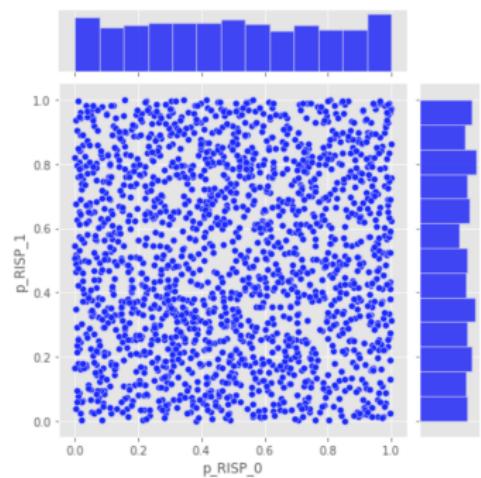


- Hall of Famer batters seemed to perform better than the rest in clutch positions
Why?

Babe versus Ruth

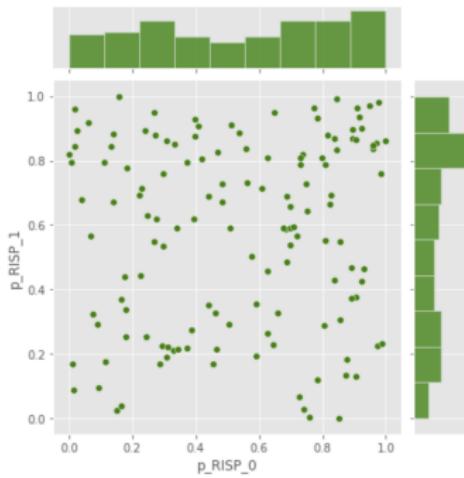
- Nate Silver proposed an experiment in his article.
If we split a player's clutch performance in even and odd year,
is there a statistical difference?

All players



$$r = 0.0903, p = 0.2942$$

HOFers



$$r = 0.0118, p = 0.4078$$

Takeaways (if you trust)

- Clutch batting ability might exists in some groups, but should be insignificant even if it does exists
- By aggressively accepting low correlation as causality:
 - Sluggers (batters good at creating long hits) might be worse in clutch
 - Batters who struck out more often might be worse in clutch
 - Batters who walks more often might be better in clutch
- Hall of famers might have a tendency to be good clutch hitters.

Future goals

- Find a better regression model which better estimates the difference between statics in and not in clutch.
- Eliminate more factors (arena, season)
- Use more advanced statistics
 - Fangraph indeed have a “clutch” statistic with a very complicated formula
 - MLB started to provide even more statistics since 2015.
 - But the older players we are interest, the less data we can get.

Outline
o

Intro
ooo

Data acquisition
ooo

Modeling clutch
oooooooo

Factors affecting clutch
ooooo

Group comparison
ooo

Conclusions
ooo●

Q & A

Thank you!