



# Baseball II



Produced by Dr. Mario | UNC STOR 538

# Pythagorean Theorem



- **Classic Pythagorean Theorem**
  - Relationship Between the Sides of a Right Triangle
  - $a^2 = b^2 + c^2$
- **What is Known: More Runs = More Wins**
- **Relationship Between Runs and Wins?**
  - Bill James' Pythagorean Method

WP = Win %  
RS = Runs Scored  
RA = Runs Allowed

$$WP \approx \frac{RS^2}{RS^2 + RA^2}$$

- **Example: Kansas City in 2014 World Series**
  - 651 Runs Scored
  - 624 Runs Allowed

$$89 \text{ Wins} \approx 162 \times \frac{651^2}{651^2 + 624^2} = 84.43$$



# Pythagorean Theorem

WP = Win %  
RS = Runs Scored  
RA = Runs Allowed

- Optimization of Relationship
  - What is the Best Choice of  $\alpha$ ?

$$WP = \frac{RS^\alpha}{RS^\alpha + RA^\alpha} + \epsilon \longrightarrow \text{Error}$$

- Minimization of Mean Absolute Deviation (MAD)
- Optimal:  $\alpha = 1.8$  (MAD=0.0199)
- Alternative Expression

$$WP = \frac{(RS/RA)^\alpha}{(RS/RA)^\alpha + 1} + \epsilon$$

- Forecasting Playoff Series Winners (2005-2016)
  - Pythagorean Method: 54.8% Accurate
  - Games Won Approach: 55% Accurate
  - Interesting Case: 2005 Nationals



# Pythagorean Theorem

WP = Win %  
RS = Runs Scored  
RA = Runs Allowed

- Useful for Valuing Players in Trades
  - Example: Cleveland Indians
  - Currently: RS=870 and RA=800
  - Trade Bing Crosby (100 Runs)
  - For Frank Sinatra (120 Runs)
  - Difference: +20 Runs
  - Before Trade:

$$WP \approx \frac{\left(\frac{870}{800}\right)^{1.82}}{\left(\frac{870}{800}\right)^{1.82} + 1} = 0.538$$

- After Trade:

$$WP \approx \frac{\left(\frac{890}{800}\right)^{1.82}}{\left(\frac{890}{800}\right)^{1.82} + 1} = 0.548$$





# Runs-Created Approach

- Motivation: Mike Trout Vs. Kris Bryant

Mike Trout and Kris Bryant 2016 Statistics

| Event                  | Trout (2016) | Bryant (2016) |
|------------------------|--------------|---------------|
| At Bats                | 549          | 603           |
| Batting Average        | .315         | .292          |
| Slugging Percentage    | .550         | .554          |
| Hits                   | 173          | 176           |
| Singles                | 107          | 99            |
| Doubles                | 32           | 35            |
| Triples                | 5            | 3             |
| Home Runs              | 29           | 39            |
| Walks + Hit by Pitcher | 127          | 93            |







# Runs-Created Approach

- **Argument**
  - Hitting Causes Good and Bad Things
  - Hits and Walks Create Scoring Opportunities
  - Better Hitter = More Scoring Opportunity
  - Relationship of Runs and {S,D,T,HR,BB,HBP}

- **Runs-Created Formula**

- Bill James (1979)
- Recall: Total Bases (TB)  
 $TB \approx S + 2D + 3T + 4HR$
- Formula:

$$RC \approx \underbrace{(H + BB + HBP)}_{\text{\# of Base Runners}} \times \frac{TB}{\underbrace{AB + BB + HBP}_{\text{Rate Players are Advancing}}}$$

H = Hit

S = Single

D = Double

T = Triple

HR = Home Run

AB = At-bat

BB = Walk

HBP = Hit-by-Pitch





# Runs-Created Approach

$y$  = Actual Runs  
 $\hat{y}$  = Predicted Runs  
 $n$  = Sample Size

- **Evaluation of Runs Created Formula**
  - Evaluated For Teams from 2010 to 2016
  - Formula for Mean Percentage Error:

$$MPE = \frac{100\%}{n} \times \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i}$$

- Based off Formula for RC, MPE = 3% (21 Runs)

- **Problem: Formula Developed Off Team Statistics**

Model Based On Teams



Predict on Players

- **Results**

| Playa and Year | Runs Created |
|----------------|--------------|
| Bryant 2016    | 129.09       |
| Trout 2016     | 134.02       |
| Cabrera 2013   | 147.54       |





# Runs-Created Approach

- **Runs Created Per Game**

- RC Flaw= Biased Toward Plate Appearances
- Observation 1: 1.8% of AB are E

$$AB - H - (0.018)AB = (0.982)AB - H$$

- Observation 2: Additional Outs Caused by GIDP, SF, SAC, and CS

$$TO = (0.982)AB - H + GIDP + SF + SAC + CS$$

- Observation 3: Sometimes 27 Outs Per Game

$$\text{Average Outs Per Game} = 26.83$$

- Observation 4: Following in Units of Game

$$\frac{TO}{26.83}$$

RC = Runs Created

AB = At-bat

E = Errors

H = Hits

TO = Total Outs

GIDP = Double-Play

SF = Sacrifice Fly

SAC = Sacrifice Bunt

CS = Caught Stealing







# Runs-Created Approach

- Runs Created Per Game
  - Final Formula for RC/G

$$\frac{RC}{G} = \frac{RC}{\frac{TO}{26.83}}$$

- Interpretation of RC/G

$$\frac{RC}{G} = \frac{\text{Runs Created by Batter}}{\text{\# of Games Worth of Outs Used by Batter}}$$

- Results Updated

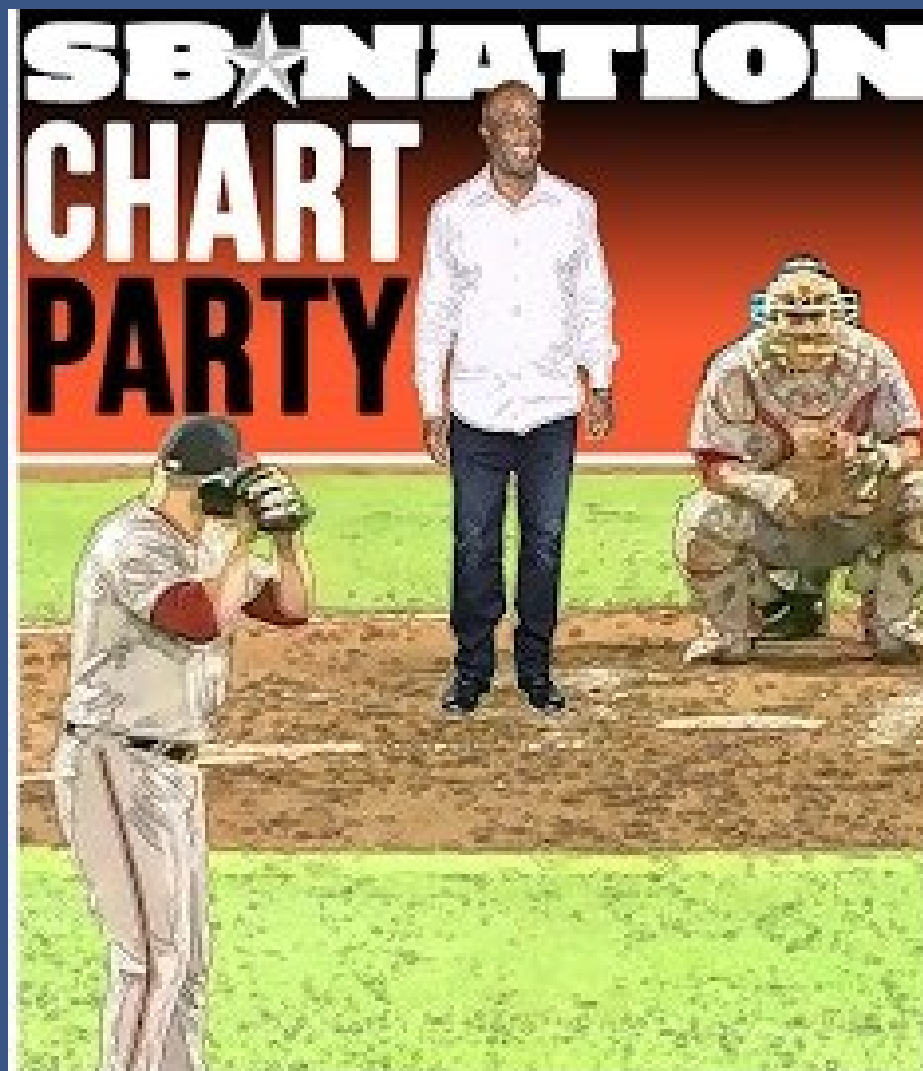
| Playa and Year | RC     | RC/G |
|----------------|--------|------|
| Bryant 2016    | 129.09 | 8.11 |
| Trout 2016     | 134.02 | 9.39 |
| Cabrera 2013   | 147.54 | 10.6 |

RC = Runs Created  
AB = At-bat  
E = Errors  
H = Hits  
TO = Total Outs  
GIDP = Double-Play  
SF = Sacrifice Fly  
SAC = Sacrifice Bunt  
CS = Caught Stealing





# America's Greatest Pastime



What if  
**Barry Bonds**  
had played  
baseball  
without  
a bat?



# America's Greatest Pastime





# Final Inspiration

Well, it took me 17 years to get 3,000 hits in baseball, and I did it in one afternoon on the golf course.

- Hank Aaron