Baseball IV

Produced by Dr. Mario | UNC STOR 390

Monte Carlo Simulation

- Recall Evaluation of Hitter Effectiveness
  - Runs Created
  - Linear Weights
  - Both Based on Team Data
  - Scaled Player Information for Prediction
- Problem: Player Hits HR 50% of Time
  - 54 RC/G Estimated by Formula (Bill James)
  - 36.8 RC/G Estimated by Linear Weights
- Definition of Monte Carlo Simulation
  - Developing a Computer Model to Repeatedly Play Out an Uncertain Situation
  - Used Across All Industries
  - Term Coined by Polish Physicist Stanislaw Ulam
  - Simple Simulation Shows Previously Discussed Player = 27 RC/G

Monte Carlo Simulation

- Monte Carlo Simulation in R
  - Theoretical Player Either Hits a Home Run or Gets an Out

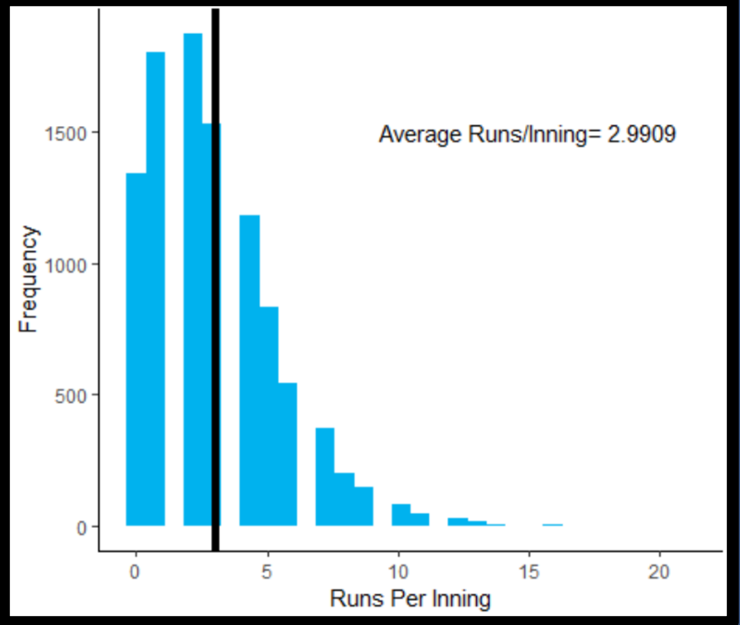
```
HR.OUT.MC=function(home.run.percent,n.Sim){
  runs.result = rep(NA,n.Sim)
  for(i in 1:n.Sim){
    runs=0
    outs=0
    while(outs<3){
      sample=runif(1)
      if(sample>home.run.percent){
        outs=outs+1
      }else{
        runs=runs+1
      }
    }
    runs.result[i]=runs
  }
  return(runs.result)
}
```

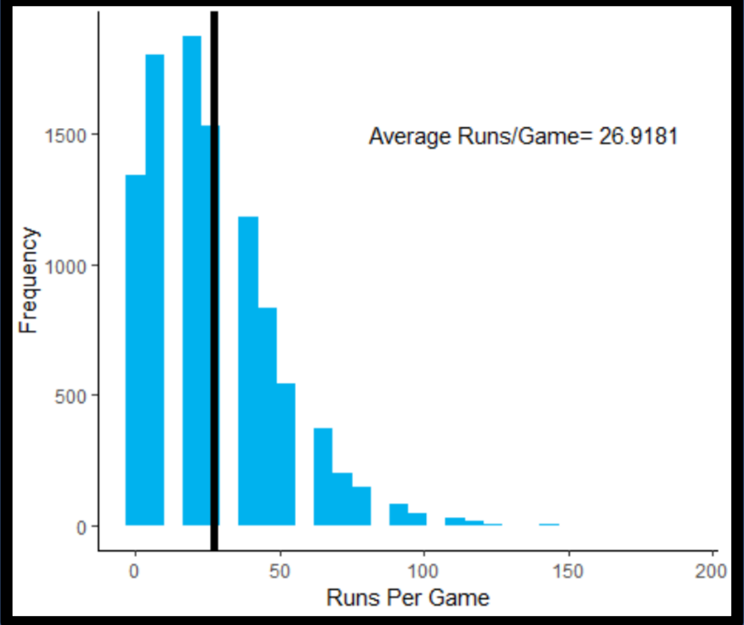
Monte Carlo Simulation

- Monte Carlo Simulation in R
  - Suppose Player Hits Home Run 50% of the Time

```
Player.5=HR.OUT.MC(0.5,10000)
Player.5=tibble(R.per.I=Player.5,
                R.per.G=Player.5*9)
```

```
head(Player.5)
A tibble: 6 x 2
R.per.I R.per.G
  <dbl>   <dbl>
      1       9
      1       9
      0       0
```





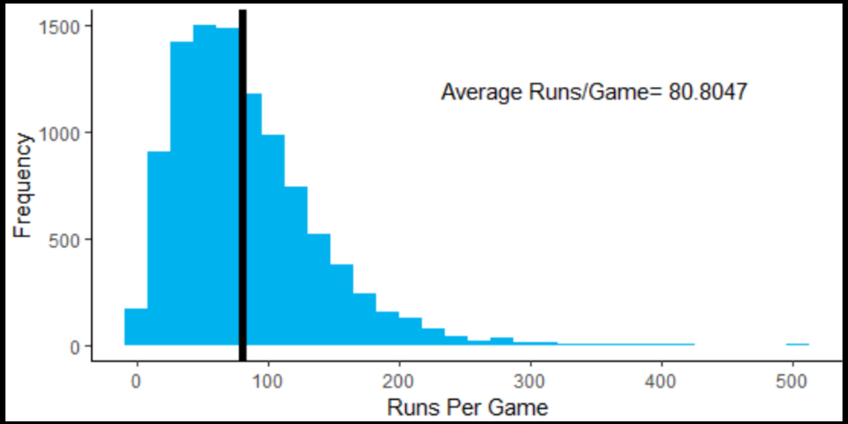
Monte Carlo Simulation

- Monte Carlo Simulation in R
  - Suppose Player Hits Home Run 75% of the Time

```
Player.75=HR.OUT.MC(0.75,10000)
Player.75=tibble(R.per.I=Player.75,
                 R.per.G=Player.75*9)
```

```
ggplot(Player.75) +
  geom_histogram(aes(x=R.per.G),fill="deepskyblue2") +
  geom_vline(xintercept=mean(Player.75$R.per.G),size=2) +
  ylab("Frequency") + xlab("Runs Per Game")+
  annotate("text", x = 350, y = 1200,size=4,
    label = paste("Average Runs/Game=",mean(Player.75$R.per.G))) +
  theme_classic()
```

```
head(Player.75)
A tibble: 6 x 2
R.per.I R.per.G
  <dbl>   <dbl>
     11      99
      7      63
     23     207
     13     117
      1       9
      3      27
```



Monte Carlo Simulation

- Simulating Runs from Team Full of Trouts
  - Possible Plate Appearances Events
  - Long List of Assumptions
    - Errors Advance All Base Runners 1 Base
    - Long Single Advances Each Runner 2 Bases
    - Short Single Advances All Runners 1 Base
    - Short Double Advances Each Runner 2 Bases
    - Long Double Scores a Runner from First
    - Etc.
  - Assign Probabilities According to Relative Frequencies of Player
  - Program for Simulation

Event
Strikeout
Walk
Hit by pitch
Error
Long single (advance 2 bases)
Medium single (score from 2nd)
Short single (advance one base)
Short double
Long double
Triple
Home run
Ground into double play
Normal ground ball
Line drive or infield fly
Long fly
Medium fly
Short fly

Monte Carlo Simulation

- Simulating Runs from Team Full of Trouts
  - Probabilities Based on Trout 2016 Statistics

Outcome	Number	Probability
Plate Appearances	681	
At Bats+ Sacrifice Hits + Sacrifice Bunts	554	
Errors	10	0.0146843
Outs (in Play)	234	0.3436123
Strikeouts	137	0.2011747
Walks	116	0.1703377
Hit by Pitch	11	0.0161527
Singles	107	0.1571219
Doubles	32	0.0469897
Triples	5	0.0073421
Home Runs	29	0.0425844

Monte Carlo Simulation

- Simulating Runs from Team Full of Ichiros
  - Probabilities of Special Cases
    - 30% of Singles are Long Singles
    - 50% of Singles are Medium Singles
    - 20% of Singles are Short Singles
    - 53.8% of Outs in Play are Ground Balls
    - 15.3% of Outs in Play are Infield Flies
    - 30.9% of Outs in Play are Fly Balls
    - Etc.
  - Result of Simulation = Within 1% of True Actual Runs Per Game
  - Specific to Trout
    - Random Number < 0.157 = Single
    - 0.157 < Random Number < (0.157+0.047) = Double
  - Goal of Simulation
    - Estimate # of Runs for Thousands of Innings
    - Average Across All Innings
    - Multiply by $\frac{26.72}{3} \approx 9$ to estimate RC/G

Monte Carlo Simulation

- **Results Under Simulation**

Player	Year	RC/G
Trout	2016	9.38
Bryant	2016	7.95
Cabrera	2013	10.24
Bonds	2004	21.02

Problem: Unusual # of Intentional Walks
Eliminating Intentional Walks: 15.98 RC/G

Monte Carlo Simulation

- Added Value of Mike Trout To LA Angels

Outcome	Number
Plate Appearances	681
At Bats+ Sacrifice Hits + Sacrifice Bunts	554
Errors	10
Outs (in Play)	234
Strikeouts	137
Walks	116
Hit by Pitch	11
Singles	107
Doubles	32
Triples	5
Home Runs	29

Trout Alone

Outcome	Number
Plate Appearances	5360
At Bats+Sacrifice Hits + Sacrifice Bunts	4962
Errors	89
Outs (in play)	2782
Strikeouts	854
Walks	355
HBP	40
Singles	848
Doubles	247
Triples	15
Home Runs	127

Without Trout

2016 Angels: 717 Runs

Sim: 626 Runs

Monte Carlo Simulation

- Added Value of Mike Trout To Average Team

Outcome	Number
Plate Appearances	6153
At Bats+Sacrifice Hits + Sacrifice Bunts	5593
Errors	101
Outs (in play)	2784
Strikeouts	1299
Walks	503
HBP	55
Singles	918
Doubles	275
Triples	29
Home Runs	187

Average Team

Outcome	Number
Plate Appearances	681
At Bats+ Sacrifice Hits + Sacrifice Bunts	554
Errors	10
Outs (in Play)	234
Strikeouts	137
Walks	116
Hit by Pitch	11
Singles	107
Doubles	32
Triples	5
Home Runs	29

Trout

How Would We Simulate Average Team + Trout?

Pitching Evaluation and Forecast

ER = Earned Run
IP = Innings

- Hypothetical Pitcher Ricky Vaughn
  - Situation 1
    - Ricky Lets 2 Batters on Base
    - Next Batter Gets Single and 1 Batter Scores
    - Ricky is Charged with 1 Earned Run
  - Situation 2
    - Ricky Lets 2 Batters on Base
    - Next Batter Hits Ball to Outfielder Who Drops the Ball
    - This Unearned Run is Not Charged to Ricky
  - Recall: ERA = Earned Run Average

$$ERA = 9 \times \frac{ER}{IP}$$

  - Ricky Gives Up 22 Earned Runs in 72 innings

$$ERA = 9 \times \frac{22}{72} = 2.75$$

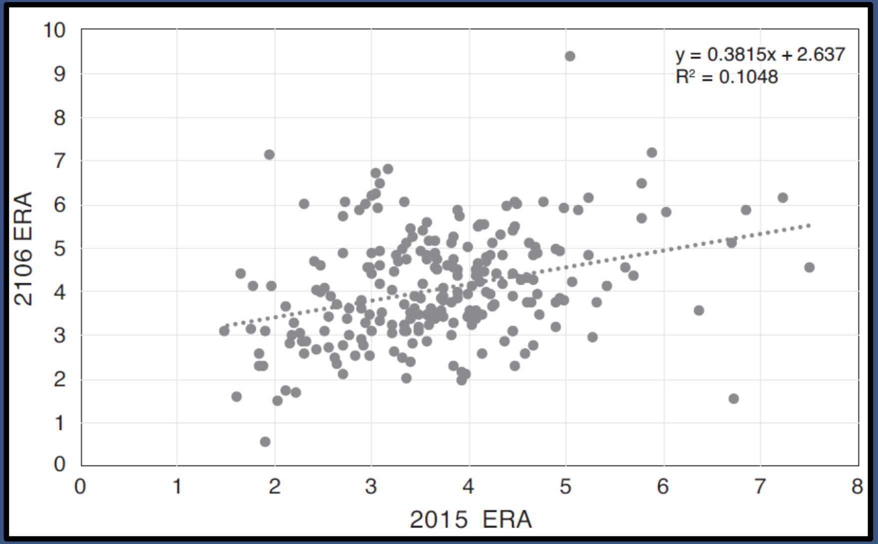
Pitching Evaluation and Forecast

ER = Earned Run
IP = Innings

- Problems with ERA
  - Influenced by Errors (Subjective)
  - Influenced by Relief Pitcher
  - Influenced by Fielding Performance

- Different Pitchers Evaluated Differently
  - Starting Pitchers = Wins
  - Relief Pitchers = Saves

- Past ERA to Predict Future ERA
  - Why Predict Future ERA?
  - Weak Relationship
  - Results Based on All Pitchers Who Pitched both Seasons



Pitching Evaluation and Forecast

- Evaluating Forecast Error
  - Mean Absolute Deviation (MAD)

$$MAD = \frac{1}{n} \times \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

  - From ERA Model, MAD = 0.93

- Additional Measures of Pitcher Effectiveness
  - Analysis by Voros McCracken (2001)
  - Fraction of Batters Faced by Pitchers That Result in Balls in Play
  - Fraction of Balls in Play That Result in Hits
  - Fraction of Batters Faced by Pitchers That Do Not Result in Balls in Play
  - Defense Independent Pitching Stats (DIPS)
    - K, BB, HBP, and HR
    - Independent of Teams Fielding Ability

y = Current ERA
$\hat{y}$ = Forecast ERA
K = Strikeout
BB = Walk
HBP = Hit by Pitch
HR = Home Run

Difficult to Predict

Pitching Evaluation and Forecast

- Defense-Independent Component ERA
  - Formula

$$DICE = C + \frac{13 \times HR + 3(BB + HBP) - 2K}{IP}$$

  - C is usually around 3.1
  - Only DIPS Involved in Formula for DICE

  - Forecast Model

$$\widehat{ERA}_t = 2.44 + 0.44 \times DICE_{t-1}$$

  - Correlation is 0.37 Compared to 0.34 when Last Year's ERA is Used
  - MAD is 0.64 Compared to 0.93 when Last Year's ERA is Used
  - Conclusion: Previous DICE is a Better Predictor of ERA than Previous ERA

K = Strikeout
BB = Walk
HBP = Hit by Pitch
HR = Home Run
IP = Inning Pitched
t = Time (Years)

America's Greatest Pastime

Final Inspiration

Politicians are like batters.
The best do their job 1/3 of the time.

-Mahatma Mario