# 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**

- **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

```r
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)
}
```
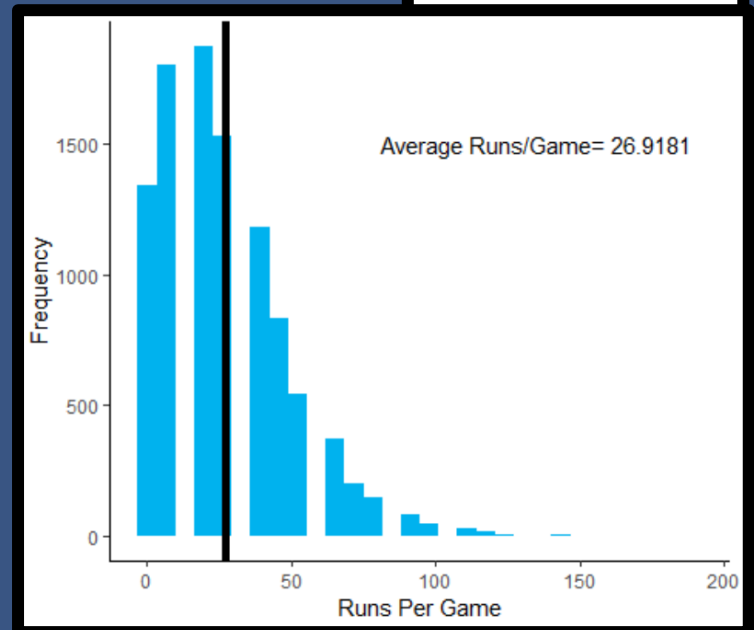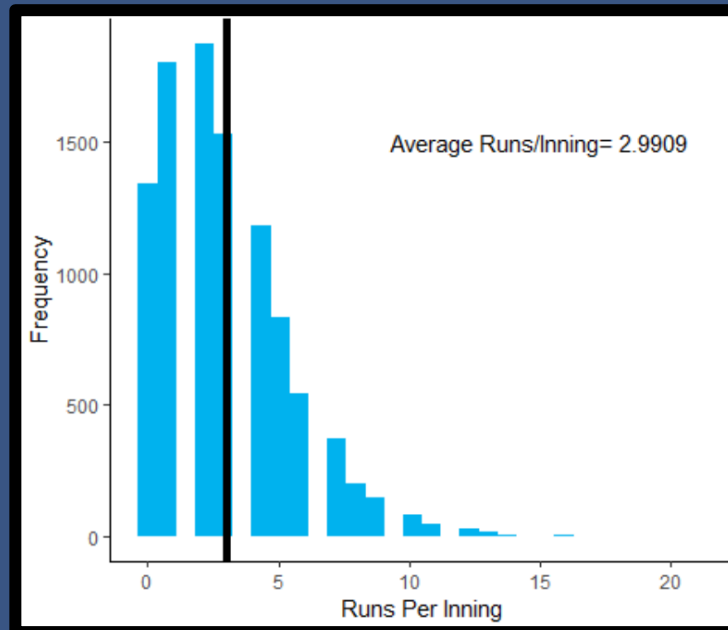
# Monte Carlo Simulation

- **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
```



Average Runs/Inning= 2.9909



Average Runs/Game= 26.9181

# 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
```



Average Runs/Game= 80.8047
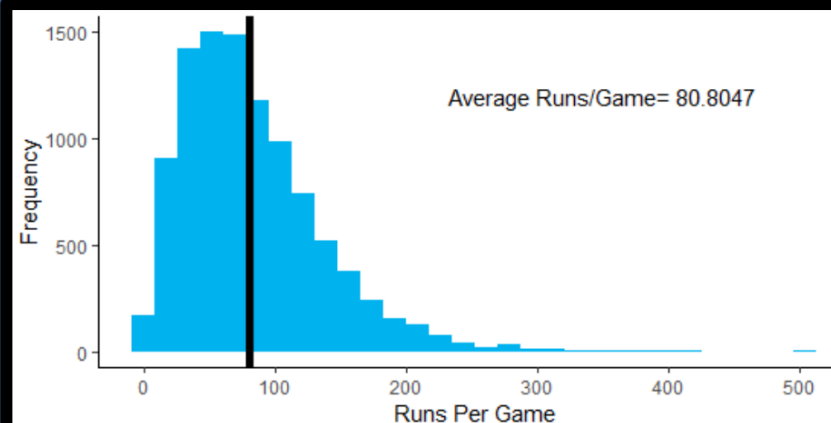
# Monte Carlo Simulation

- **Simulating Runs from Team Full of Ichiros**
  - **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 Ichiros**
  - **Probabilities Based on Ichiro 2004 Statistics**

|  | Number | Probability |
|---|---|---|
| Plate Appearances | 762 | |
| At Bats +Sac. Hits + Sac. Bunts | 709 | |
| Errors | 13 | 0.0170604 |
| Outs (in play) | 371 | 0.4868766 |
| Strikeouts | 63 | 0.0826772 |
| BB | 49 | 0.0643045 |
| HBP | 4 | 0.0052493 |
| Singles | 225 | 0.2952756 |
| 2B | 24 | 0.0314961 |
| 3B | 5 | 0.0065617 |
| HR | 8 | 0.0104987 |

# 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 Ichiro**
    - **Random Number < 0.295 = Single**
    - **0.295 < Random Number < (0.295+0.487) = Out (In-Play)**
  - **Goal of Simulation**
    - **Estimate # of Runs for Thousands of Innings**
    - **Average Across All Innings**
    - **Multiply by $\dfrac{26.72}{3} \approx 9$ to estimate RC/G**

# Monte Carlo Simulation

- **Results Under Simulation**

| Player | Year | RC/G |
|--------|------|------|
| Ichiro | 2004 | 6.92 |
| Nomar | 1997 | 5.91 |
| Bonds | 2004 | 21.02 |

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

# Monte Carlo Simulation

- **Added Value of Albert Pujols Measured by Runs**

| Outcome | Number |
|---|---|
| Plate Appearances | 5591 |
| At Bats + Sac. Hits + Sac. Bunts | 5095 |
| Errors | 92 |
| Outs (in Play) | 2824 |
| Strikeouts | 872 |
| BB | 439 |
| HPB | 57 |
| Singles | 887 |
| 2B | 259 |
| 3B | 26 |
| HR | 135 |

| Outcome | Number |
|---|---|
| Plate Appearances | 634 |
| At Bats +Sac. Hits + Sac. Bunts | 538 |
| Errors | 10 |
| Outs (in play) | 301 |
| Strikeouts | 50 |
| BB | 92 |
| HBP | 4 |
| Singles | 94 |
| 2B | 33 |
| 3B | 1 |
| HR | 49 |

Average Team With

Team Without

Pujols Alone

Sim: 706 Runs

783 to 853 Runs

| Outcome | Number |
|---|---|
| Plate Appearances | 6236.27 |
| At Bats +Sac. Hits + Sac. Bunts | 5658.03 |
| Errors | 102 |
| Outs (in play) | 3027.23 |
| Strikeouts | 1026.37 |
| BB | 528.23 |
| HBP | 50 |
| Singles | 986.67 |
| 2B | 304.5 |
| 3B | 31.73 |
| HR | 179.53 |

# 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**

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

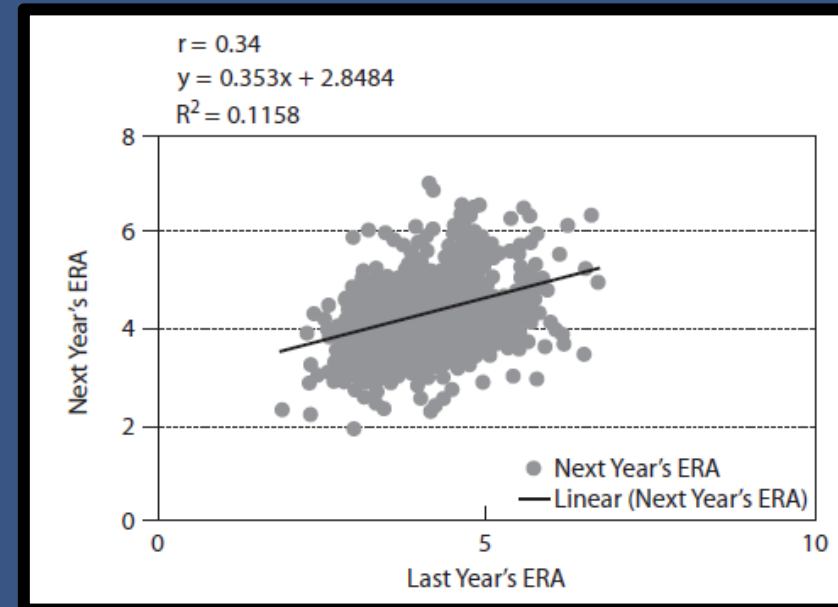  - **Ricky Gives Up 22 Earned Runs in 72 innings**

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

# 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**
  - **Low Linear Correlation**
  - **Results Based on Pitchers with More than 10 Innings**



$r = 0.34$
$y = 0.353x + 2.8484$
$R^2 = 0.1158$

- Next Year's ERA
— Linear (Next Year's ERA)

Last Year's ERA
Next Year's ERA

# Pitching Evaluation and Forecast

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

- **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 = 0.68**

- **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**

# Pitching Evaluation and Forecast

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

- **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 = 0.68**

- **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**

# Pitching Evaluation and Forecast

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

- **Defense-Independent Component ERA**
  - **Formula**

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

  - **Only DIPS Involved in Formula for DICE**
  - **Forecast Model**

$$ERA_t = 1.975 + 0.56 \times DICE_{t-1}$$

  - **Correlation is 0.44 Compared to 0.34 when Last Year's ERA is Used**
  - **MAD is 0.51 Compared to 0.68 when Last Year's ERA is Used**
  - **Conclusion: Previous DICE is a Better Predictor of ERA than Previous ERA**

- **Holy Grail of Mathletics = Forecasting Performance**

America's Greatest Pastime

WORST CELEBRITY FIRST PITCHES!

# Final Inspiration

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

-Mahatma Mario