2023 World Baseball Classic

Monte Carlo Simulation

Judy Chan / chuting3@illinois.edu
Cindy Yang / tyy3@illinois.edu

Background

2023 WBC just ended in March, and we looked into the system, finding that **tournament format** and **pitching counts** are two important variables.

Table of contents

Experiment 1

- Pitching Count

Hypothesis

The more pitches the starting pitcher throws, the higher the probability of winning.

What are the pitch count rules in the World Baseball Classic?

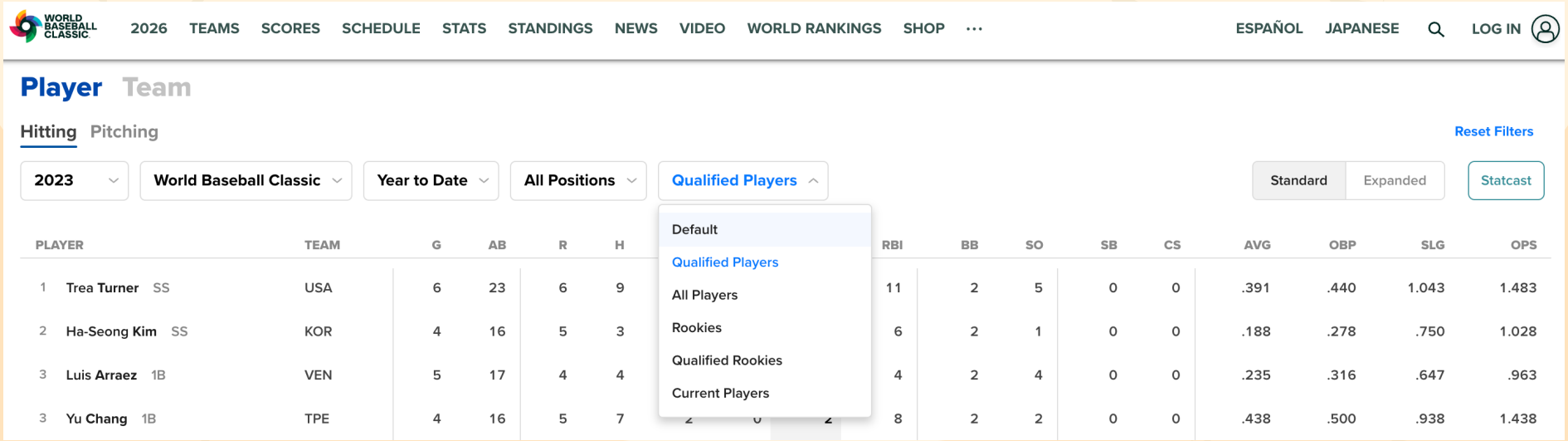
There are restrictions on how many pitches a player can throw and how often they can be used.

The pitch count regulations are different for each stage of the tournament:

- 65 pitches per game in the first round
- 80 pitches per game in the quarterfinals
- 95 pitches per game in the semifinals and final

Assumptions

- 3 pitchers and 9 batters per team; no substitutions
- Pitcher importance based on pitch count; equal importance for all batters
- Game outcomes determined by combined pitcher and batter performance
- Data from 2023 WBC top eight countries; excludes non-participating players



Phase 1 - Design

Determining Victory: $$\text{Total Performance} = \text{Total Pitching Score} + \text{Total Hitting Score}$$

Random Variables:

1. Player selection:
   - Randomly select 3 pitchers from data/pitchers.csv
   - Randomly select 9 batters from data/batters.csv

pitchers.csv

Project_WBC_Simulation

CSV

	player	team	W	L	ERA	G	GS	CG	SHO	SV	SVO	IP	H	R	ER	HR	HB	BB	SO	WHIP	AVG
1	Mitch Neunborn	AUS	0	0	4.76	3	0	0	0	0	1	5.2	3	3	3	0	0	3	8	1.06	0.15
2	Kyle Glogoski	AUS	1	0	0	1	1	0	0	0	0	2.2	1	0	0	0	0	3	5	1.5	0.111
3	Sam Holland	AUS	0	0	4.5	4	0	0	0	0	0	2	3	1	1	0	1	1	3	2	0.375
4	Jack O' Loughlin	AUS	0	0	1.93	2	2	0	0	0	0	4.2	1	1	1	0	0	0	3	0.21	0.067
5	ToddVan Steensel	AUS	0	0	0	2	0	0	0	0	0	3	2	0	0	0	0	0	3	0.67	0.182

batters.csv

	player	team	G	AB	R	H	2B	3B	HR	RBI	BB	SO	SB	CS	AVG	OBP	SLG	OPS
1	Robbie Glendinning	AUS	5	20	4	6	0	0	2	6	3	6	1	0	0.3	0.391	0.6	0.991
2	Alex Hall	AUS	5	22	3	5	1	1	2	5	2	5	0	0	0.227	0.292	0.636	0.928
3	Tim Kennelly	AUS	5	16	5	4	1	0	1	2	3	5	0	0	0.25	0.4	0.5	0.9
4	RobbiePerkins	AUS	5	16	3	3	0	0	1	3	1	2	0	0	0.188	0.278	0.375	0.653
5	RixonWingrove	AUS	5	16	2	6	1	0	1	7	2	4	0	0	0.375	0.444	0.625	1.069

The meaning of each performance indicator is provided in data/variables.txt

Phase 1 - Design

Random Variables:

2. Generating normally distributed simulation data:
   - Based on season performance → Fluctuating player performance

pitcher_records_AUS.csv

	player	team	Random_ERA	Random_WHIP	Random_BAA
1	KyleGlogoski	AUS	0.0	1.5197065737916335	0.11532109616844564
2	KyleGlogoski	AUS	0.0	1.4611221263257885	0.12184735457725372
3	KyleGlogoski	AUS	0.0	1.4208801519279866	0.10620321159141712
4	KyleGlogoski	AUS	0.0	1.4029047949420044	0.1114502530861664
5	KyleGlogoski	AUS	0.0	1.459749041080547	0.12324964762914689
6	KyleGlogoski	AUS	0.2723601775910374	1.5380869339772414	0.11545344121606899

For example, Kyle's season ERA (Earned Run Average) is 0.3, if there're 100 games simulated,
The program will generate 100 random batting averages (x1 to x100), which follow normal distribution:
mu = 0.3, std = 0.2 (Different indicators have varying standard deviations due to real-world conditions)

$$\frac{1}{100}\sum_{i=1}^{100} x_i = 0.3$$

Phase 1 - Design

Random Variables:

3. Randomly allocate the number of pitches for the three pitchers and use them as weights when calculating performance scores:

$$\text{Pitching Score per pitcher} = (0.3 \times \text{ERA}) + (0.25 \times \text{WHIP}) + (0.15 \times \text{BAA}) + (0.1 \times \text{IP}) + (0.2 \times \text{K})$$

$$\text{Total Pitching Score} = (\frac{\text{p1 pitch count}}{\text{total pitch count}}) \times \text{p1} + (\frac{\text{p2 pitch count}}{\text{total pitch count}}) \times \text{p2} + (\frac{\text{p3 pitch count}}{\text{total pitch count}}) \times \text{p3}$$

And also, calculating hitting performance for 9 randomly selected batters:

$$\text{Hitting Score per batter} = (0.25 \times \text{BA}) + (0.3 \times \text{OPS}) + (0.15 \times \text{RBI}) + (0.1 \times \text{BB}) + (0.1 \times \text{SO}) + (0.1 \times \text{SB})$$

$$\text{Total Hitting Score} = \text{B1} + \text{B2} + \text{B3} + \text{B4} + \text{B5} + \text{B6} + \text{B7} + \text{B8} + \text{B9}$$

4. Total Pitch Count: 120 ~ 200

Phase 2 - Validation

Comparing Simulation Results to 2023 WBC Rankings:

We will compare our simulation results to the actual rankings of the 2023 WBC to see if they align with reality.

The top eight teams in the competition are as follows:

1. Japan (JPN)
2. United States (USA)
3. Mexico (MEX)
4. Cuba (CUB)
5. Venezuela (VEN)
6. Puerto Rico (PUR)
7. Australia (AUS)
8. Italy (ITA)

By analyzing our simulation results, we can determine if our model accurately predicts the performance of these teams.

Phase 3 - Experiment

Two teams with similar strengths

By observing the conditional probability (i.e., the probability of Team A winning when their starting pitcher has a higher pitch count), it can be inferred that when the starting pitcher has a higher pitch count, the team is more likely to win.

```
1. AUS 2. CUB 3. ITA 4. JPN 5. MEX 6. PUR 7. USA 8. VEN
Enter the number of country A: 1
Enter the number of country B: 8
Enter the count of simulation: 1000
```

sim	t1-Country	t1-p1_cnt	t1-p1_cnt_%	t1-result	t2-Country	t2-p1_cnt	t2-p1_cnt_%	t2-result
1	AUS	86	0.61	win	VEN	71	0.51	lose
2	AUS	53	0.3	lose	VEN	129	0.72	win
3	AUS	60	0.39	win	VEN	78	0.5	lose
4	AUS	72	0.4	lose	VEN	125	0.69	win
5	AUS	53	0.37	lose	VEN	95	0.67	win

```
Summary Statistics
simulation times: 1000
Team 1                                          Team 2
P(A|B)    win_times     lose_times    win_rate=P(B)   P(C|D)    win_times     lose_times    win_rate=P(D)
0.66      452           548           0.45            0.62      548           452           0.55
```

P(A|B) = P(p1 using more pitches | Team A wins)

Phase 3 - Experiment

Two teams with a significant difference in strength

When observing the conditional probability, it can be noticed that for the stronger team, the number of pitches thrown by the starting pitcher does not have a significant impact on the win rate. However, for the weaker team, a large proportion of their wins occur when their starting pitcher throws more pitches than the starting pitcher of the stronger team.

```
Please choose two countries to match up:
1. AUS 2. CUB 3. ITA 4. JPN 5. MEX 6. PUR 7. USA 8. VEN
Enter the number of country A: 4
Enter the number of country B: 3
Enter the count of simulation: 1000
```

sim	t1-Country	t1-p1_cnt	t1-p1_cnt_%	t1-result		t2-Country	t2-p1_cnt	t2-p1_cnt_%	t2-result
1	JPN	111	0.57	win		ITA	119	0.61	lose
2	JPN	106	0.69	win		ITA	65	0.42	lose
3	JPN	91	0.49	win		ITA	89	0.48	lose
4	JPN	89	0.49	win		ITA	72	0.4	lose
5	JPN	81	0.55	win		ITA	98	0.67	lose

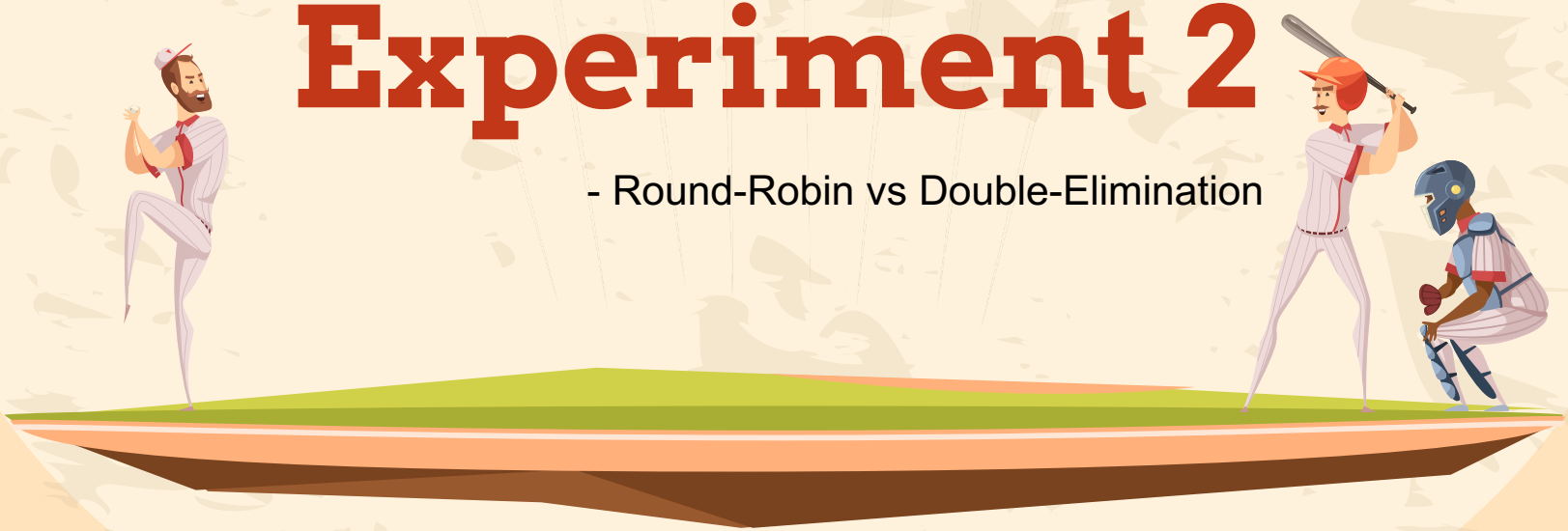
Summary Statistics

simulation times: 1000

Team 1				Team 2			
P(A|B)	win_times	lose_times	win_rate=P(B)	P(C|D)	win_times	lose_times	win_rate=P(D)
0.51	916	84	0.92	0.85	84	916	0.08

Experiment 2

- Round-Robin vs Double-Elimination



Background

2006

Round-Robin

2009

Double-Elimination

2013

Round-Robin
Mix
Double-Elimination

2017

Round-Robin

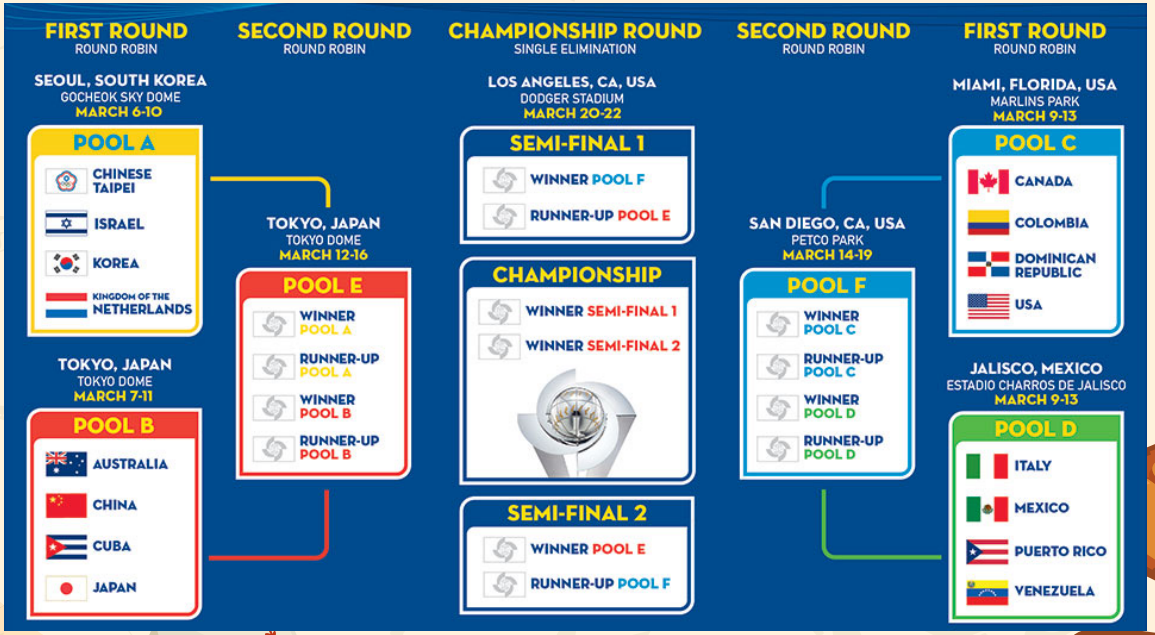
2023

Round-Robin

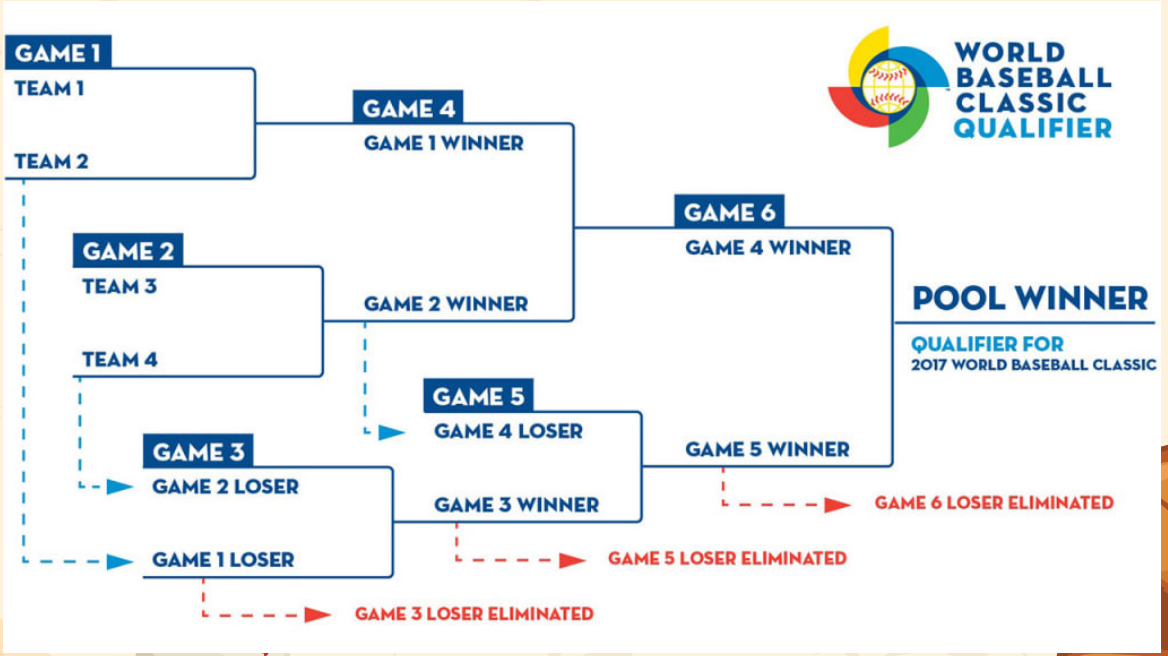
Hypothesis

The round-robin format is more advantageous for teams with stable performance than the double-elimination format.

Round-Robin


FIRST ROUND
ROUND ROBIN
SECOND ROUND
ROUND ROBIN
CHAMPIONSHIP ROUND
SINGLE ELIMINATION
SECOND ROUND
ROUND ROBIN
FIRST ROUND
ROUND ROBIN
SEOUL, SOUTH KOREA
GOCHEOK SKY DOME
MARCH 6-10
POOL A
CHINESE TAIPEI
ISRAEL
KOREA
KINGDOM OF THE NETHERLANDS
TOKYO, JAPAN
TOKYO DOME
MARCH 7-11
POOL B
AUSTRALIA
CHINA
CUBA
JAPAN
TOKYO, JAPAN
TOKYO DOME
MARCH 12-16
POOL E
WINNER POOL A
RUNNER-UP POOL A
WINNER POOL B
RUNNER-UP POOL B
LOS ANGELES, CA, USA
DODGER STADIUM
MARCH 20-22
SEMI-FINAL 1
WINNER POOL F
RUNNER-UP POOL E
CHAMPIONSHIP
WINNER SEMI-FINAL 1
WINNER SEMI-FINAL 2
SEMI-FINAL 2
WINNER POOL E
RUNNER-UP POOL F
SAN DIEGO, CA, USA
PETCO PARK
MARCH 14-19
POOL F
WINNER POOL C
RUNNER-UP POOL C
WINNER POOL D
RUNNER-UP POOL D
MIAMI, FLORIDA, USA
MARLINS PARK
MARCH 9-13
POOL C
CANADA
COLOMBIA
DOMINICAN REPUBLIC
USA
JALISCO, MEXICO
ESTADIO CHARROS DE JALISCO
MARCH 9-13
POOL D
ITALY
MEXICO
PUERTO RICO
VENEZUELA

Double-Elimination



Assumptions

- 3 pitchers and 9 batters per team; no substitutions
- 3 pitchers would be equally important, as we do not consider pitch counts here
- Game outcomes are only determined by combined pitcher-batter performance and the competition system
- Data from 2023 WBC top eight countries; excludes non-participating players
- Stable Performance can be represented by the gaming result of 2023 WBC

Phase 1 - Design

Determining Victory:

$$\text{Total Performance} = \text{Total Pitching Score} + \text{Total Hitting Score}$$

Random Variables:

1. Player selection:
   - Randomly select 3 pitchers from data/pitchers.csv
   - Randomly select 9 batters from data/batters.csv

pitchers.csv

Project_WBC_Simulation

CSV

	player	team	W	L	ERA	G	GS	CG	SHO	SV	SVO	IP	H	R	ER	HR	HB	BB	SO	WHIP	AVG
1	Mitch Neunborn	AUS	0	0	4.76	3	0	0	0	0	1	5.2	3	3	3	0	0	3	8	1.06	0.15
2	Kyle Glogoski	AUS	1	0	0	1	1	0	0	0	0	2.2	1	0	0	0	0	3	5	1.5	0.111
3	Sam Holland	AUS	0	0	4.5	4	0	0	0	0	0	2	3	1	1	0	1	1	3	2	0.375
4	Jack O' Loughlin	AUS	0	0	1.93	2	2	0	0	0	0	4.2	1	1	1	0	0	0	3	0.21	0.067
5	ToddVan Steensel	AUS	0	0	0	2	0	0	0	0	0	3	2	0	0	0	0	0	3	0.67	0.182

batters.csv

	player	team	G	AB	R	H	2B	3B	HR	RBI	BB	SO	SB	CS	AVG	OBP	SLG	OPS
1	Robbie Glendinning	AUS	5	20	4	6	0	0	2	6	3	6	1	0	0.3	0.391	0.6	0.991
2	Alex Hall	AUS	5	22	3	5	1	1	2	5	2	5	0	0	0.227	0.292	0.636	0.928
3	Tim Kennelly	AUS	5	16	5	4	1	0	1	2	3	5	0	0	0.25	0.4	0.5	0.9
4	RobbiePerkins	AUS	5	16	3	3	0	0	1	3	1	2	0	0	0.188	0.278	0.375	0.653
5	RixonWingrove	AUS	5	16	2	6	1	0	1	7	2	4	0	0	0.375	0.444	0.625	1.069

The meaning of each performance indicator is provided in data/variables.txt

Phase 1 - Design

Random Variables:

2. Pool generation:
   - Round-Robin Simulation: 2 pools of four teams
   - Double-Elimination Simulation: 4 pools of two teams

```
 -----Round Robin Format-----
First round:
{'pool_A': {'MEX': 1, 'CUB': 1, 'AUS': 1, 'JPN': 3}, 'pool_B': {'ITA': 0, 'VEN': 2, 'USA': 2, 'PUR': 2}}
Countries get into the second round:  ['JPN', 'MEX', 'PUR', 'VEN']
```

```
 -----Double Elimination Format-----

 14 Matches:
<Match left=<Participant ITA> right=<Participant USA> winner=<Participant USA> loser=<Participant ITA>>
<Match left=<Participant MEX> right=<Participant VEN> winner=<Participant MEX> loser=<Participant VEN>>
<Match left=<Participant CUB> right=<Participant PUR> winner=<Participant CUB> loser=<Participant PUR>>
<Match left=<Participant JPN> right=<Participant AUS> winner=<Participant JPN> loser=<Participant AUS>>
```

Phase 1 - Design

Random Variables:

3. Run rate:
   - Round-Robin Simulation: happens when there's a tie, we randomized the value of total runs allowed

$$\text{Run Rate} = \text{Total Runs Allowed}/\text{Total Number of Outs}$$

Phase 2 - Validation

Comparing Simulation Results to 2023 WBC Rankings:

We will compare our simulation results to the actual rankings of the 2023 WBC to see if they align with reality.

The top eight teams in the competition are as follows:

1. Japan (JPN)
2. United States (USA)
3. Mexico (MEX)
4. Cuba (CUB)
5. Venezuela (VEN)
6. Puerto Rico (PUR)
7. Australia (AUS)
8. Italy (ITA)

By analyzing our simulation results, we can determine if our model accurately predicts the performance of these teams.

SECOND ROUND

QUARTERFINAL 1

3 AUSTRALIA
POOL B RUNNER-UP

vs

4 CUBA
POOL A WINNER

CUBA ADVANCES

QUARTERFINAL 3

9 USA
POOL C RUNNER-UP

vs

7 VENEZUELA
POOL D WINNER

USA ADVANCES

CHAMPIONSHIP ROUND

SEMIFINAL 1

2 CUBA

vs

14 USA

USA ADVANCES

SEMIFINAL 2

5 MEXICO

vs

6 JAPAN

JAPAN ADVANCES

CHAMPIONSHIP

2 USA

vs

3 JAPAN

JAPAN
2023 WORLD BASEBALL CLASSIC CHAMPIONS

SECOND ROUND

QUARTERFINAL 2

3 ITALY
POOL A RUNNER-UP

vs

9 JAPAN
POOL B WINNER

JAPAN ADVANCES

QUARTERFINAL 4

4 PUERTO RICO
POOL D RUNNER-UP

vs

5 MEXICO
POOL C WINNER

MEXICO ADVANCES

Phase 3 - Experiment

```
 -----Round Robin Format-----
First round:
{'pool_A': {'MEX': 1, 'PUR': 2, 'VEN': 3, 'ITA': 0}, 'pool_B': {'AUS': 1, 'CUB': 0, 'USA': 2, 'JPN': 3}}
Countries get into the second round:  ['VEN', 'PUR', 'JPN', 'USA']

Second round:
{'pool_A': {'USA': 1, 'PUR': 2, 'JPN': 3, 'VEN': 0}}
Countries get into the final round:  ['JPN', 'PUR']
JPN wins the 2023 WBC!
```

Phase 3 - Experiment

```
 -----Double Elimination Format-----

 14 Matches:
<Match left=<Participant MEX> right=<Participant VEN> winner=<Participant MEX> loser=<Participant VEN>>
<Match left=<Participant ITA> right=<Participant CUB> winner=<Participant ITA> loser=<Participant CUB>>
<Match left=<Participant PUR> right=<Participant JPN> winner=<Participant JPN> loser=<Participant PUR>>
<Match left=<Participant USA> right=<Participant AUS> winner=<Participant AUS> loser=<Participant USA>>
<Match left=<Participant MEX> right=<Participant AUS> winner=<Participant MEX> loser=<Participant AUS>>
<Match left=<Participant ITA> right=<Participant JPN> winner=<Participant JPN> loser=<Participant ITA>>
<Match left=<Participant USA> right=<Participant VEN> winner=<Participant VEN> loser=<Participant USA>>
<Match left=<Participant PUR> right=<Participant CUB> winner=<Participant PUR> loser=<Participant CUB>>
<Match left=<Participant MEX> right=<Participant JPN> winner=<Participant JPN> loser=<Participant MEX>>
<Match left=<Participant AUS> right=<Participant PUR> winner=<Participant PUR> loser=<Participant AUS>>
<Match left=<Participant ITA> right=<Participant VEN> winner=<Participant VEN> loser=<Participant ITA>>
<Match left=<Participant PUR> right=<Participant VEN> winner=<Participant PUR> loser=<Participant VEN>>
<Match left=<Participant MEX> right=<Participant PUR> winner=<Participant MEX> loser=<Participant PUR>>
<Match left=<Participant JPN> right=<Participant MEX> winner=<Participant JPN> loser=<Participant MEX>>

Semi Final:  ['JPN', 'MEX', 'PUR', 'VEN']
Final:  ['MEX', 'JPN']
JPN wins the 2023 WBC!
```

Phase 3 - Experiment

By observing the probability of each country entering Semi-Final and Final, it can be inferred that round-robin and double-elimination does not differentiate the results a lot.

-----Round Robin Format-----

Country	#Semi	Semi-Final Prob.	#Final	Final Prob.
AUS	203	0.20	26	0.03
CUB	589	0.59	260	0.26
ITA	201	0.20	23	0.02
JPN	987	0.99	968	0.97
MEX	529	0.53	192	0.19
PUR	311	0.31	60	0.06
USA	579	0.58	225	0.23
VEN	601	0.60	246	0.25

-----Double Elimination Format-----

Country	#Semi	Semi-Final Prob.	#Final	Final Prob.
AUS	275	0.28	34	0.03
CUB	541	0.54	302	0.30
ITA	237	0.24	10	0.01
JPN	994	0.99	966	0.97
MEX	522	0.52	215	0.21
PUR	289	0.29	30	0.03
USA	582	0.58	270	0.27
VEN	560	0.56	173	0.17

Thank You!

References

- World Baseball Classic Data
- Word Baseball Classic Wikipedia
- Round-Robin Tournament
- Double-Elimination Tournament
- Double-Elimination package