Lecture 10: Probability and Expectation in Sports

Stanford University
Stats 60
Will Hartog

April 21, 2025

Lecture Outline

1. Numbers and Sports
2. Win Probability in Football
3. Expected Batting Average in Baseball
4. Expected Points in Football
5. Expected Runs in Baseball

Why does this lecture exist?

Based off of material from Stats 100: Mathematics of Sports. Previously taught by me, currently taught yearly by Gene Kim!

Agree or Disagree: Sports are inherently numbers-driven.

Thesis: Fans, media, and athletes understand and talk about sports with numbers and data.

"They don't ask how, they ask how many."

Example: What's in a GOAT?

Serena Williams: 23 Grand Slam titles, 14 Grand Slam Doubles titles with sister Venus, 20 years between first and last Slam final appearance, 186 weeks spent at world #1.

Tom Brady: 7 Super Bowl rings with 2 teams, 80,214 passing yards, 649 TD passes, 15 Pro Bowls, 251 career wins, 3 scandals.

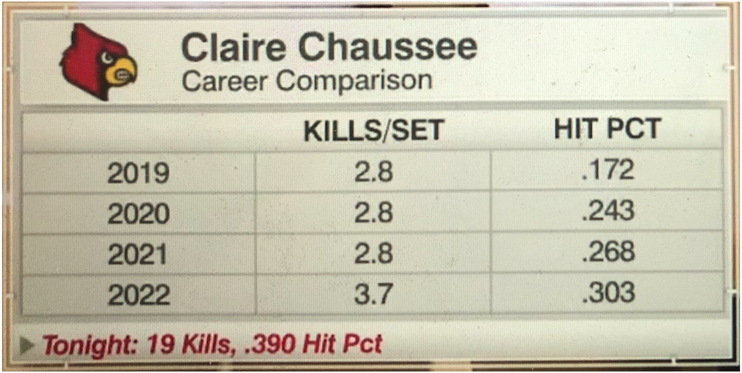
Example: What's in a GOAT?

Katie Ledecky: 6 individual Olympic gold medals, 42 total medals in international competition, 14 world records in 3 events, Stanford Class of 2021.

LeBron James: 4 NBA championships with 3 different teams, 50,365 career points, 21 time all-star, $163.7 million in *Space Jam 2* box office.

Example: Sports Graphics

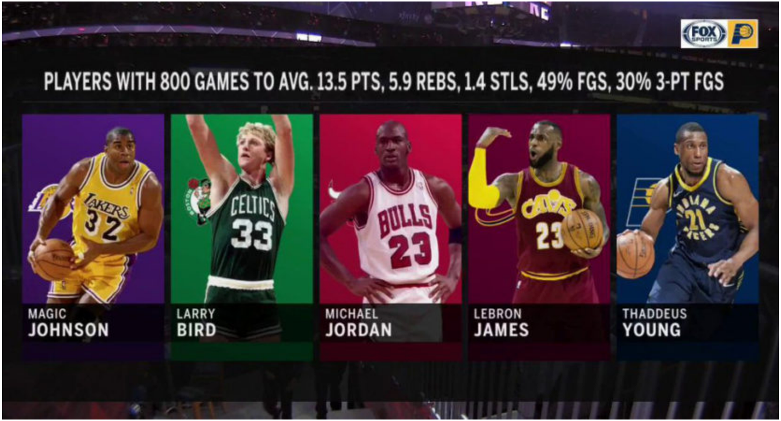
The media frequently uses graphics with athlete statistics, usually selected to make a specific point.


Claire Chaussee
Career Comparison

	KILLS/SET	HIT PCT
2019	2.8	.172
2020	2.8	.243
2021	2.8	.268
2022	3.7	.303

Tonight: 19 Kills, .390 Hit Pct

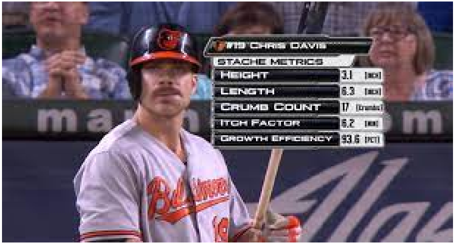
Example: Sports Graphics



Example: Sports Graphics

Sports graphics gone too far

Sports graphics gone too far



Example: Determining Outcomes

The outcome of a sporting event is literally based on numbers.

- LSU beat Iowa **82 - 59** to win the 2025 Women's NCAA Basketball Tournament.
- Tiger Woods won the 2019 Men's Masters Tournament with a score of **13 under par**, compared to runner-ups at **12 under par**.
- Stanford beat Calfornia **2** games to **1** to win their weekend series on the Farm.

Don't just take my word for it: The greats know it's just math

Where do Math and Statistics come in?

Frederick Mosteller (1916-2006): “It is easy to lie with statistics; it is easier to lie without them.”

- Statistics is the art of learning from the data.
- Out of context numbers can be misleading.
- We want to determine what is luck (noise), and what is skill (signal).
- Statistics provides mathematical models to interpret our data and quantify uncertainty.

Probabilities in Sports

Q: What are some examples of *probabilities* in sports?

Win Probability

We often think about having an estimate for *win probability* model for each moment in time.

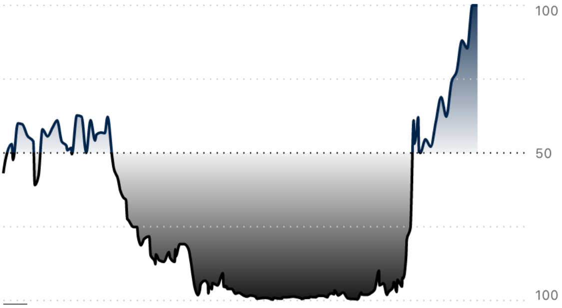


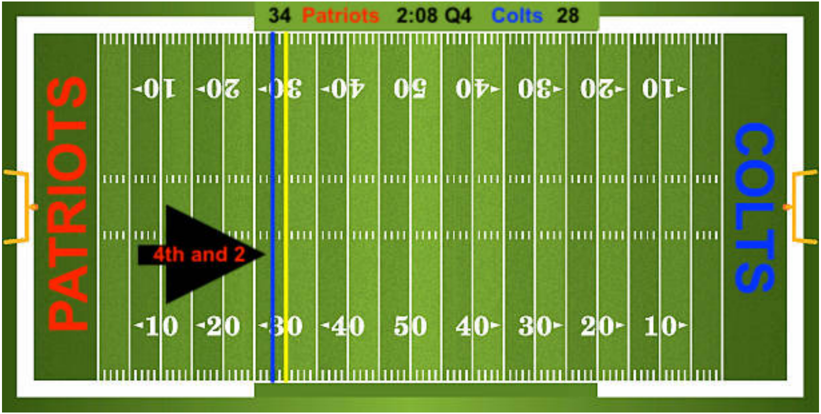
Figure: Win probability for Super Bowl LI between the Patriots and Falcons.

Computing Win Probability

- What factors (context) do we need to compute win probability in football?
- Usual answer:
    - Down (1st, 2nd, 3rd, 4th)
    - Distance
    - Yard Line
    - Time left
- We can compute win probability from a historical average of what percent of teams won a game in a similar sitution.
- In general there are more sophisticated ways to do this.

Decision Making

Example: November 15th, 2009. NFL Week 10 Sunday Night Football.



Q: What's the right play?

Quantifying decision value

We need to compute the probability of winning for each possible decision (go for it or punt)

- The Patriots have a 94.6% to win with a conversion, and a 50.9% chance with a failed conversion.

$$WP_{GO-FOR-IT} = 0.946 \cdot z + 0.509 \cdot (1 - z) \approx 0.741,$$

  where $z \approx 0.53$ is the league-wide probability of converting.
- A punt on average gives the Colts the ball at their own 33-yard line, which gives

$$WP_{PUNT} = 0.736.$$

- We can find the value of z we'd need to justify going for it, which is 52%, just below league average.

Q: Which decision do you support now?

What actually happened?.

Modern Example

What do you do when the estimated probabilities are so close?

Twitter Bot for Detroit Lions @ SF 49ers, NFC Conference Championship

Down 3, 4th & 3, 30 yards from opponent end zone

Qtr 4, 07:32 | Timeouts: Off 3, Def 2

			Win % if	
	Win %	**Success %**[1]	**Fail**	**Succeed**
Go for it	**28**	53	15	40
Field goal attempt	**26**	71	13	32

[1] Likelihood of converting on 4th down or of making field goal

Source: @ben_bot_baldwin

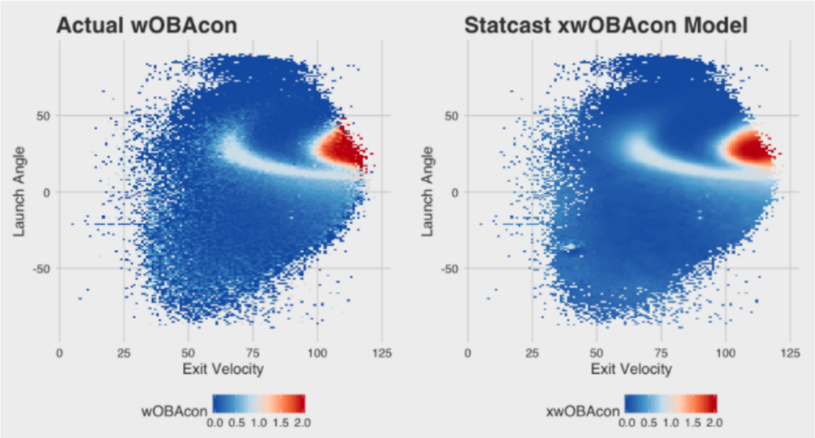
Q: Which decision do you support? Is it because of the model (or an issue with the model), or is it because of football logic?

Probabilities for Individual Plays

- Sometimes it may make sense to assign probabilities to individual plays.
- Can help determine if a player/team got lucky.
- Helps determine team and player quality.
- We will look at outcomes in baseball, namely whether a batted ball results in a hit.
- In 2015 Statcast was introduced, and today we have very useful semi-publicly available data on how hard the baseball was hit, how far it went up, launch angle, etc.

Idea: Can we quantify the quality of a hit by how they hit the ball?

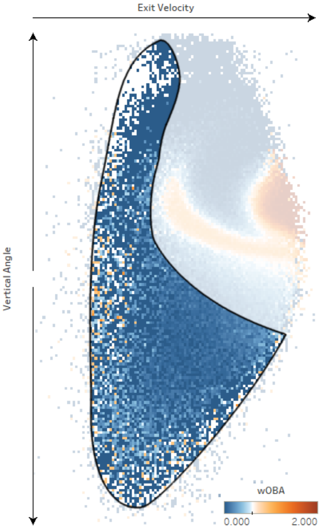
Statcast data



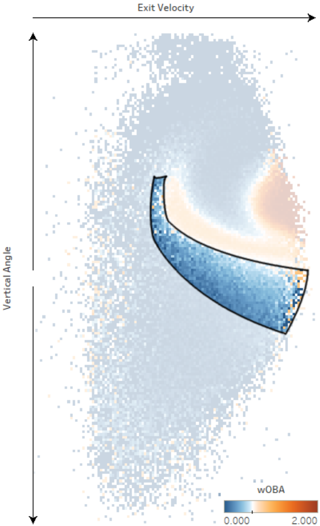
Statcast to the rescue

- You can build a model based on the launch angle and exit velocity!
- 100 mph is roughly the upper limit of how fast the pitchers can throw, but while hitting you add your own physical energy and the collision is roughly elastic. So you get exit velocity above 100.
- Color coding is how good the outcome was.

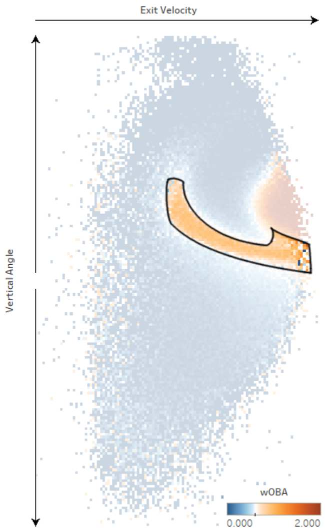
Dribble balls: most weakly hit ground balls



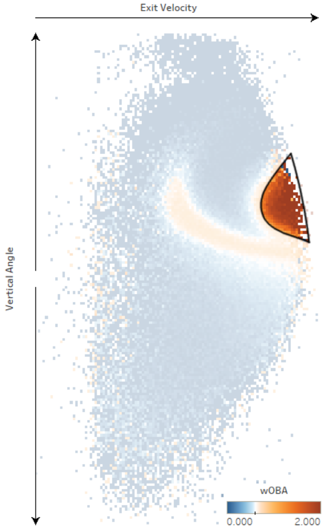
Ground balls



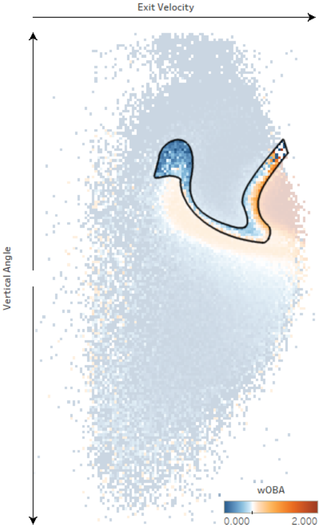
Low drive: mostly singles



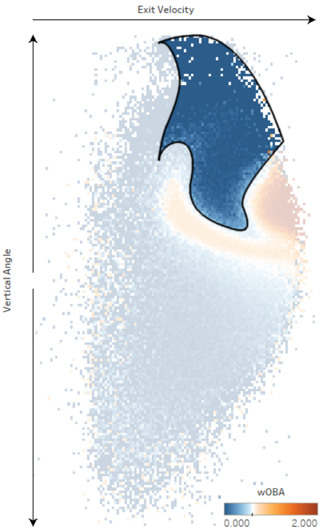
High drive: Best of all, mostly home runs



Fly ball



Pop up



Estimating the Probability

- Once a batter hit the ball, we get the values for its exit velocity and launch angle.
- Then we can look at the k closest historical batted balls to estimate the probability that it resulted in each outcome type.
- This method is called k-Nearest Neighbors, because you compare the closest k "neighbors". This is an example of a statistical *model* for estimating a parameter, in this case, the probability of a hit.
- For example, if you take the $k = 10$ closest points and 90% neighbors lead to HR and 10% lead to out, you conclude that the pitch has a 90% chance of resulting in HR and 10% chance of resulting in out.

Brainstorming Expected Probabilities in Other Sports

Question: Find your 1-2 “nearest neighbors” and come up with at least two other sports (besides baseball and football) in which you could computed the probability of some outcome. What context and data would you need to estimate this probability?

Completion Probability in the NFL

Uses ‘Next Gen Stats’, a recent development, we compute completion probability using location of defenders, both near target and quarterback, time to throw, etc.

Many different ways to interpret, but very useful to have.

While Next Gen Stats weren’t available in 2008, likely a low completion probability for the famous David Tyree “helmet catch”.

Shot Probability in Basketball

Similar idea, uses tracking data (WNBA and NBA) like distance to defender, location on floor, time elapsed in shot clock, etc.

Expected Goals (xG) in Soccer

Becoming much more ubiquitous, uses all kinds of tracking data.

BREAK

George Costanza Explains Analytics

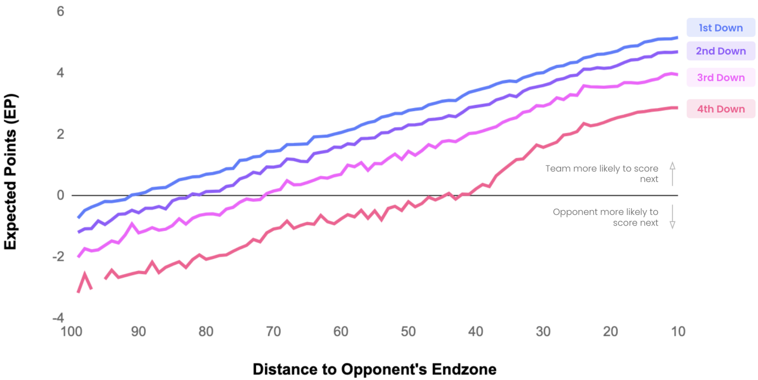
Dan Campbell masterclass:

Fourth down 1

Fourth down 2

Expected Points in Football

Similar to win probability, we can compute with historical data the expected points scored from a given state.



This graphic excludes distance to the next first down from the state space.

How to Compute

Say we have n historical data points in a state s. We could estimate directly:

$$\text{Expected Points from State } s = EP_s = \frac{1}{n}\sum_{i=1}^{n} Pts_i^s,$$

where Pts_i^s is the next scoring event in data point i at state s.

This scoring event could be the other team's score (in which case Pts_i^s is negative).

Definition: Expectation is the average value an outcome can take.

Our value for expected points is an *estimate* of the true average.

Expected Points Added

Given a specific play, we can compute the expected points before and after and the difference is expected points added.

NWE 30	10	14	Stephen Gostkowski kicks off 67 yards, returned by Domenik Hixon for 14 yards (tackle by Raymond Ventrone)	0.000	-0.060
NYG 17	10	14	Eli Manning pass complete short right to Amani Toomer for 11 yards (tackle by Rodney Harrison)	-0.060	0.810
NYG 28	10	14	Eli Manning pass incomplete short middle intended for Plaxico Burress	0.810	0.260
NYG 28	10	14	Eli Manning pass incomplete short left intended for Plaxico Burress	0.260	-0.430
NYG 28	10	14	Eli Manning pass complete short left to Amani Toomer for 9 yards (tackle by Rodney Harrison)	-0.430	-0.910
NYG 37	10	14	Brandon Jacobs up the middle for 2 yards (tackle by Vince Wilfork and Richard Seymour)	-0.910	1.530
NYG 39	10	14	Eli Manning right tackle for 5 yards (tackle by Adalius Thomas)	1.530	1.660
NYG 39	10	14	Timeout #1 by New York Giants	1.530	1.660
NYG 44	10	14	Eli Manning pass incomplete deep right intended for David Tyree	1.660	0.960
NYG 44	10	14	Eli Manning pass complete deep middle to David Tyree for 32 yards (tackle by Rodney Harrison)	0.960	3.970
NYG 44	10	14	Timeout #2 by New York Giants	0.960	3.970
NWE 24	10	14	Eli Manning sacked by Adalius Thomas for -1 yards	3.970	3.300
NWE 24	10	14	Timeout #3 by New York Giants	3.970	3.300
NWE 25	10	14	Eli Manning pass incomplete short left intended for David Tyree (defended by Brandon Meriweather)	3.300	2.610
NWE 25	10	14	Eli Manning pass complete short right to Steve Smith for 12 yards (tackle by Brandon Meriweather)	2.610	4.710
NWE 13	16	14	Eli Manning pass complete short left to Plaxico Burress for 13 yards, touchdown	4.710	7.000
NWE 2	17	14	Lawrence Tynes kicks extra point good	0.000	0.000

Expected points before and after for the New York Giants' game-winning drive in Super Bowl XLII.

Which Plays Were Crucial

David Tyree's helmet catch was worth $3.970 - 0.960 = 3.01$ expected points added.

However, there were many high EPA plays on this drive.

Player Evaluation with EPA

We can also evaluate quarterbacks with EPA per dropback (DB) and per play.

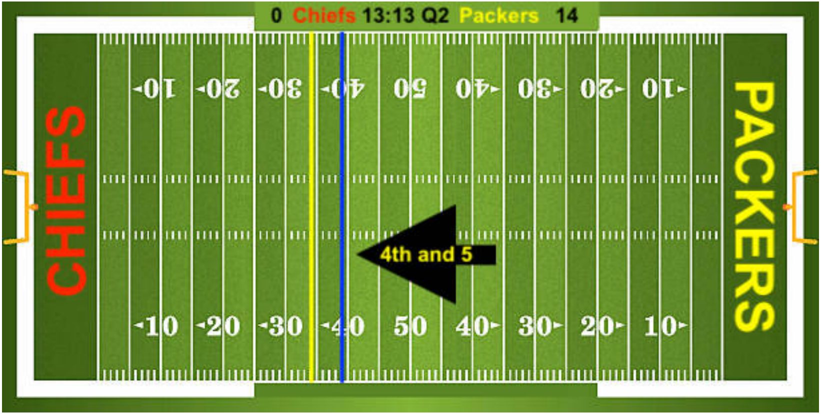
			EPA		EPA Components vs League Average						
			Total	/ DB ▼	Rushing	Sacks	Inc	INTs	Air Yards	YAC	Penalties
1	Brock Purdy	SF '23	173	0.33	0.01	0.01	0.03	-0.02	0.16	0.09	0
2	Rayne Prescott	DAL '23	175	0.24	0.01	0.01	0.03	0.02	0.06	0.03	0.02
3	Joshua Allen	BUF '23	137	0.18	0.06	0.06	0.02	0.01	0.02	0	-0.02
4	Tua Tagovail...	MIA '23	115	0.18	-0.03	0.03	0.01	-0.01	0.08	0.07	0
5	Jordan Love	GB '23	104	0.16	-0.01	0.04	-0.01	0.02	0.05	0.03	0
6	Jake Browning	CIN '23	41	0.14	0.01	0	0.04	-0.02	-0.02	0.1	0
7	Lamar Jackson	BAL '23	88	0.13	0.02	0	0.04	0.03	0.03	-0.03	0
8	Jalen Hurts	PHI '23	87	0.12	0.01	0.04	0.06	0	0.03	-0.06	0

44	Zachary Wilson	NYJ '23	-96	-0.21
45	Dorian Thomps...	CLE '23	-29	-0.22
46	Daniel Jones	NYG '23	-56	-0.24
47	Trevor Siemian	NYJ '23	-52	-0.29
48	Bailey Zappe	NE '23	-87	-0.34

2023 season (NFL Stats).

Fourth Down Decision Making Redux

Example: September 28th, 2015. NFL Week 3 Monday Night Football.



Q: What's the right play?

Quantifying decision value

Three options:

- Go for it (estimated 40% success rate on 4th and 5)
- Field goal (estimated 55% success rate from 39-yard line)
- Punt (expected starting position between 10- and 11- yard lines)

We can use our expected points model to evaluate the value of each decision.

Go for it

- For simplicity, we can assume the Packers have a 1st and 10 from the 34-yard line if they convert.
- If they fail, we'll assume an incomplete pass, so the Chiefs get the ball on their own 39-yard line.
- EP for a success is +3.28, EP for the Chiefs on a failure is +1.51.

$$EP_{GO-FOR-IT} = (3.28) \cdot x + (-1.51) \cdot (1 - x) \approx +0.41$$

where $x \approx 0.4$ is the probability of a successful conversion.

Attempt a field goal

- A success gives 3 points, minus the EP of the Chiefs after the kickoff, $\approx +0.4$, so $+2.6$ overall.
- A failure gives the Chiefs the ball at their own 47-yard line, worth $+2.02$ EP for the Chiefs.

$$EP_{FGA} = 2.6 \cdot y + (-2.02) \cdot (1 - y) \approx +0.32,$$

where $y \approx 0.55$ is the probability of a sucessful field goal.

Punt

To punt is a sad decision to make, but it is sometimes wise.

- We'll have to account for the different possibilities of a punt:
  - Touchback
  - Fair catch
  - Return
- From this we can estimate expected opponent field position $EOFP$ (about the 11-yard line), and then assign the value of a punt to their EP at that field position, which is -0.2.

$$EP_{PUNT} = -EP_{EOFP} = +0.2$$

This is a simplification, it is better but harder to get probabilities of a punt going to each yard line, then weight over those EP's.

Which do we pick?

- Comparing the expected point values for each decision:

  $$EP_{GO-FOR-IT} = +0.41, EP_{FGA} = +0.32, EP_{PUNT} = +0.2,$$

- **Q:** Which decision do you prefer now?

- What actually happened?
- **Q:** Does this result invalidate the decision?

Expected Runs in Baseball

- Say we want to compute expected runs in a given inning in baseball.
- What context do we need?
- Which bases are occupied, how many outs

Expected Run Matrix

Base Runners				2010-2015		
1B	2B	3B		0 outs	1 outs	2 outs
__	__	__		0.481	0.254	0.098
1B	__	__		0.859	0.509	0.224
__	2B	__		1.100	0.664	0.319
1B	2B	__		1.437	0.884	0.429
__	__	3B		1.350	0.950	0.353
1B	__	3B		1.784	1.130	0.478
__	2B	3B		1.964	1.376	0.580
1B	2B	3B		2.292	1.541	0.752

Figure: Avg runs scored from that config to inning end

Q: Which is the most valuable game state? Does that make sense intuitively?

Context-dependent Value

We can assign a value to each play by how much the expected runs change, including any runs that score.

- What is the value of a walk at the beginning of an inning (0 outs, no runners)?
- The run expectancy with 0 outs, 0 runners is 0.481
- The run expectancy with 0 outs, runner on 1B is 0.859
- That walk contributed $0.859 - 0.481 = 0.378$ runs
- The value of a home run in the same situation is 1, since the run expectancy before and after is 0.481, but one run scored

Averaging over States

- We can find the value of a given type of hit, e.g. walk, by averaging over the states.
- To find the appropriate weights, we need the frequency of each state.

Base Runners			2010-2015		
1B	2B	3B	0 outs	1 outs	2 outs
__	__	__	0.244	0.175	0.139
1B	__	__	0.059	0.070	0.071
__	2B	__	0.015	0.026	0.033
1B	2B	__	0.014	0.025	0.031
__	__	3B	0.002	0.009	0.014
1B	__	3B	0.005	0.011	0.016
__	2B	3B	0.003	0.007	0.008
1B	2B	3B	0.004	0.009	0.011

Figure: State frequency

2013 estimates for run values

Estimated Run Values for the 22 Most Common Events, 2013

Event	N	*Freq*	$\bar{r}$	$\hat{\beta}$
Home run	4,661	0.025	1.54	1.37
Triple	772	0.004	0.62	1.02
Double	8,185	0.044	0.40	0.75
Field error	1,516	0.008	0.20	0.49
Single	28,448	0.154	0.22	0.44
Hit by pitch	1,536	0.008	0.01	0.31
Walk	13,622	0.074	0.02	0.30
Intent walk	1,018	0.005	0.01	0.18
Sac fly	1,204	0.006	1.01	−0.00
Sac bunt	1,382	0.007	0.04	−0.10
Groundout	35,171	0.190	0.02	−0.20
Flyout	23,080	0.125	0.00	−0.23
Lineout	9,493	0.051	0.00	−0.23
Strikeout	36,573	0.197	0.00	−0.25
Popout	8,877	0.048	0.00	−0.26
Forceout	3,946	0.021	0.07	−0.31
Grounded into DP	3,731	0.020	0.03	−0.75

The rightmost column $\hat{\beta}$ is the estimated average value for each type of play.

Weighted On-Base Average (wOBA)

- wOBA is a metric that uses changes to the expected run matrix to determine the value of different events (Tango, 2007).

- wOBA is designed so that the league average wOBA is equal to the league average OBP, putting the former numbers on the familiar scale of the latter.

wOBA (2020)

$$wOBA \approx \frac{0.7 \cdot BB + 0.883 \cdot 1B + 1.24 \cdot 2B + 1.56 \cdot 3B + 1.98 \cdot HR}{PA}$$

This is approximate as certain plate appearances (PA) are not counted, and walks (BB) are broken up into walks and hit-by-pitch.

Check out Fangraphs website for wOBA coefficient values from 1871 to present. The weights are changed yearly.

Is wOBA better than OPS? It has been presented as an improvement.

Combining wOBA and Expected Hit Probability

- Using Statcast predictions for probability of each type of outcome, we can assign an expected WOBA to each batted ball.
- xwOBA can be computed using these ideas based off of the quality of batted ball, not the outcome.

MLB's Expected Statistics
Expected Stats Help in Other Ways

Wrapping Up

- We've seen a number of ways to estimate probabilities and expectations in sports.
- They can help us make decisions, evaluate players, and understand the game better.
- Can these kinds of ideas help outside of sports?

Question: With your group from before, brainstorm "expected" metrics for your daily life. Give the context for the metric, and say how you might use it.