SABERMETRICS 101

AN INTRODUCTION TO BASEBALL STATISTICS

HISTORY

1858-1970

- Box score invented in 1858 by Henry Chadwick
- Batting average is king
- Earnshaw Cook publishes *Percentage Baseball* in 1964
  - First sabermetric work to gain national media attention
  - Criticized by both statisticians and baseball people
  - Today we have shown a lot of his ideas were wrong

1971-1990

- In 1971 Bill James, Pete Palmer and Dick Cramer founded the Society for American Baseball Research (SABR)
- In 1980 Bill James invented the term “sabermetrics”
- Davey Johnson wrote FORTRAN programs in the early 1970s to try and convince his manager to let him bat second in the lineup. In 1984, as a manager, he had employees write statistical programs for him
- Craig Wright is credited as the first front office employee to have the title Sabermetrician, working for the Rangers in the 1980s

1991-PRESENT

- Oakland A's and the Moneyball Era
  - Sandy Alderson began using statistics to search for undervalued players in the early 1990s
  - Billy Beane continued these strategies when he became GM in 1997
  - Billy Beane and the A's statistical methods drew national attention and recognition when Michael Lewis published *Moneyball* in 2003
- Today, every MLB team employs statisticians and computer scientists to do their own sabermetric analysis
- Pitchf/x debuted in 2006 and was installed in every stadium
  - Used 3 separate cameras to measure the trajectory, speed, spin, break and location of every pitch
- Statcast (including TrackMan) replaced Pitchf/x and is now installed in every MLB ballpark
  - Uses doppler radio and HD video to measure both the ball and every player on the field

TRADITIONAL METRICS

THE OK, THE BAD, AND THE UGLY

PREREQUISITES

- A **Plate Appearance (PA)** occurs every time a player steps to the plate and is either put out or reaches base (exception: replaced by pinch hitter with a 2-strike count and PH strikes out)
- An **At Bat (AB)** is a plate appearance that does not end in a walk, hit by pitch, catcher's interference, sacrifice bunt, sacrifice fly, catcher's interference
- A **Hit (H)** occurs when a batter reaches first base after hitting the ball into fair territory, without the benefit of an error or fielder's choice
- **Home Run (HR)**, **Single (1B), Double (2B), Triple (3B)** are kinds of hits
- **Walk (BB), Intentional Walk (IBB), Hit By Pitch (HBP or HP), Sacrifice Bunt (SB), Sacrifice Fly (SF), Sacrifice Hit (SH = SB + SF), Catcher's Interference (CI), Fielder's Choice (FC), Error (E)** are ways of reaching base that aren't hits

BATTING AVERAGE (BA)

- Formula: H/AB = (Hits) / (At Bats)
- The standard number by which a hitter's ability is measured
- Does not penalize or reward a player for drawing walks
- Does not directly rely on the situation (runners on base, number of outs, inning)
- Does not reward the player for extra-base hits (all hits are counted the same)
- Key values:
  - .200 is considered atrocious
  - .300 is traditionally considered excellent
  - .400 hasn't been done in a season since Ted Williams

RUNS BATTED IN (RBI)

- A batter is credited one RBI for each runner that scores as a result of his at bat, provided:
  - The run doesn't score as the result of an error
  - The batter doesn't ground into forced double-play
  - The batter doesn't ground into half of what should have been a double-play (missed throw at first base)
- Important statistic because of the Triple Crown, awarded when a player has the highest batting average, RBIs and home runs in a given year
- Not a great measure of player ability as it is highly dependent on the rest of the team

WINS (W)

- Awarded to the pitcher who pitched the last half-inning before the winning team took the lead for the final time
  - Except if the pitcher is a starting pitcher and he has not lasted 5 innings
  - Or if it is a relief pitcher, who, in the judgment of the official scorer, was ineffective in a brief appearance
- It's a pitching statistic that relies on the offense
- Terrible for comparing pitchers
- 20 wins in a season is extraordinary, 25 hasn't happened since 1990
- Corresponding stat is Losses (L)
  - Loss is awarded to the pitcher who pitched the last half-inning before the losing team started losing for the last time
  - OR the pitcher who put the go-ahead run on base before getting taken out of the game

EARNED RUN AVERAGE (ERA)

- Formula: 9 * (ER/IP) = 9 * (Earned Runs)/(Innings Pitched)
- If no errors and no passed balls occurred in an inning, all runs that inning are earned
- If you're interested in the complexities of earned/unearned runs, read the rule book
- ERA is meant to give an idea of how many runs a pitcher would allow in a complete game
- A pitcher is credited with earned runs after he leaves the game if he was the pitcher to put the runners on
- Not great for evaluating pitchers as a great defense will help lower a pitcher's ERA, and, for relief pitchers, it doesn't take into account letting someone else's earned runs score

FIELDING PERCENTAGE

- Formula: (Putouts + Assists)/(Putouts + Assists + Errors)
- Does not take into account range
  - A better fielder who reaches more balls might make more errors than the worse fielder who can't reach the ball in order to make an error

BASERUNNING STATS

- Stolen base (SB)
- Caught stealing (CS)
- SB% = SB/(SB + CS)
- Defensive Indifference – doesn't count for an SB or CS

MODERN BUT STILL TRADITIONAL METRICS

THE GOOD

ON BASE PERCENTAGE (OBP)

- Formula (H + BB + HBP)/(AB + BB + HBP + SF) = (Hits+ Walks + Hit-By-Pitches)/(~Plate Appearances)
- Better measure of how often a hitter reaches base on his own merits
  - Gives credit for walks and getting hit by pitch
  - Decreases for reaching due to error and fielder's choice
- Doesn't take into account power (all hits are counted the same)
- Does give credit for intentional walks
- Range is .000 to 1.000 like batting average
- Single-season record is held by Barry Bonds, set in 2004 with an OBP of .609
- Ted Williams holds the career record with an OBP of .482

SLUGGING PERCENTAGE (SLG)

- Formula: (1B + 2*2B + 3*3B + 4*HR)/AB = (Total bases)/(At bats)
- Gives credit for extra-base hits, weighting each base as equally more important than the last (doubles count for twice singles, triples count for three times a single)
- Range is .000 (no hits in *x* at bats) to 4.000 (*x* home runs in *x* at bats)
- Barry Bonds holds the record for single-season SLG with .863 in 2001
- 2016 average slugging percentage was .417

ON BASE PLUS SLUGGING (OPS)

- Formula: OBP + SLG = (On base percentage) + (Slugging percentage)
- Turns out to be a good overall measure of offensive production as it credits getting on base and hitting for power roughly equally
- A great OPS is over .900 (the league leader is usually right around 1.000)
- Was once considered a more advanced (or sabermetric) metric, but is solidly accepted traditionally now, as it has been included on Topps baseball cards since 2004
- It can be criticized as an overall offensive metric for two main reasons:
  - It adds OBP and SLG which have similar but not identical denominators so it's not pretty
  - It doesn't take into account base running, situational performance or base stealing

SAVE (S OR SV)

- Started being recorded in some form in 1952, popularized by *Sporting News* in 1960, and became an official statistic in 1969
- According to the rule book, a save is awarded to a pitcher if and only if:
  - He is the finishing pitcher in a game won by his team;
  - He is not the winning pitcher;
  - He is credited with at least ⅓ of an inning pitched; and
  - He satisfies one of the following conditions:
    - He enters the game with a lead of no more than three runs and pitches for at least one inning
    - He enters the game, regardless of the count, with the potential tying run either on base, at bat or on deck
    - He pitches for at least three innings.
- Hold (H) is a similar (but not official) statistic, awarded to a pitcher who satisfies the above conditions but does not finish the game

WALKS + HITS PER INNING PITCHED (WHIP)

- Formula is in the name: (BB + H)/IP = (Walks + Hits)/(Innings Pitched)
- Invented in 1979 but has become mainstream
- Like OPS for hitters, WHIP gives a good idea of a pitcher's overall effectiveness
- League leader will usually be at or below 1.00
  - Record for a single season is held by Pedro Martinez with a WHIP of 0.7373
  - Only three pitchers have had a career WHIP under 1.00, Mariano Rivera's was exactly 1.00

ADVANCED METRICS

AN ATTEMPT TO PERFECTLY QUANTIFY EACH PLAYER IN EVERY WAY

WHY ADVANCED METRICS?

- The game changes over time
- Good players play for bad teams, and bad players play for good teams
- Every ballpark has different dimensions and is in a different environment
- Every hitter faces different pitches from various pitchers in diverse situations throughout a season
- Sometimes it's more important to perform well than others
- Everyone would like to have just one number to define how good a player is
- Despite the amount of data we have, sometimes there's still luck

wOBA (WEIGHTED ON BASE AVERAGE)

- The formula FanGraphs uses is *wOBA = (0.690×uBB + 0.722×HBP + 0.888×1B + 1.271×2B + 1.616×3B +2.101×HR) / (AB + BB – IBB + SF + HBP)*
- The constants change each year as the run-value of each type of hit changes
- The constants are calculated using linear weights, which is outside the scope of this presentation. The basic idea is you make a matrix of the possible situations that can occur and the number of runs that result in each situation based on each event you're interested in. You then divide that number of runs by the total number of that event that occurred in the season to get the run expectancy of that event on average
- Scaling is then applied to these constants to put wOBA on the same scale as OBP. Scaling takes into account the expected run value of outs (negative) and then finds the ratio of league wOBA (with unscaled values) to league OBP to find the multiplier for each constant to get the values in the above formula

BASE RUNS AND RUNS CREATED

- Two different statistics that attempt to answer the question – how many runs **should** be scored given offensive events
- Runs created in its most basic form usually predicts a team's actual runs scored within 5%
- Basic form: OBP * SLG * AB ~= ((H+BB) * TB)/(AB + BB) ~= TB * OBP
- Most versions take into account other offensive stats like HBP, SB, IBB
- Base Runs is significantly more accurate at predicting actual runs, and uses linear weights (like wOBA) to better estimate the run-producing potential of each offensive stat.

WINS ABOVE REPLACEMENT (WAR)

- The number of additional wins a player's team would achieve above the number of wins they would be expected to win with a replacement-level player
- Formula is usually something like: *WAR = (Batting Runs + Base Running Runs +Fielding Runs + Positional Adjustment + League Adjustment +Replacement Runs) / (Runs Per Win)*
- Everyday players are usually in the 2-4 range, all star players are 4-6, and the MVP will have 6+
- Calculated differently for hitters and pitchers, but on the same scale
- Different sources calculate it slightly differently so you can't compare numbers you see on different websites
- Main criticism is that it isn't great for comparing players across eras, as baseball players have gotten collectively better and the gap between best and worst player has narrowed

FIELDING INDEPENDENT PITCHING (FIP)

- Formula: $\frac{13*HR + 3*(BB+HBP) - 2*K}{IP}$ + {FIPCONSTANT}
- The constant is only to bring FIP onto the same scale as ERA
- Only uses statistics that are independent of a team's defense and luck
- Assumes league-average results on balls in play
- xFIP is the same as FIP but instead of actual HR allowed, uses the expected number of home runs a pitcher would have given up given the league-average fly ball to home run rate

STATCAST METRICS

- Exit velocity, generated velocity, launch angle, distance, pitch location, spin rate, perceived velocity
- xBA - uses Statcast's hit probability to determine likelihood of a hit given similar batted balls
- xSLG – uses hit probability + TB likelihood
- xwOBA – same as xSLG but uses the wOBA formula
- Acceleration, throwing velocity, first step, catch probability, exchange time
- Pop time for catchers

FUTURE TOPICS/DISCUSSION

- Clutch score
- Win-probability added
- Win-shares
- Plate discipline statistics
- Ultimate Zone Rating (UZR)
- Pitch-framing
- Simulations and fantasy baseball
- General baseball strategy (bullpen/bench management, win probability, how sabermetrics has affected player management and baseball operations)

FURTHER READING

- FanGraphs – my favorite baseball blog – fangraphs.com
  - Great sabermetric articles every day
  - Lots of glossary entries for basic and advanced metrics
  - Search functionality to see all these stats for players and teams
- Baseballsavant Statcast Search – baseballsavant.mlb.com/statcast_search
  - Very searchable database of (almost?) all Statcast data
  - Leaderboards for common metrics
- *Moneyball* by Michael Lewis
- *Smart Baseball* by Keith Law
- *The Book* by Tom Tango, Mitchel Lichtman and Andrew Dolphin