# Intro

* **Sabermetrics** = “the search for objective knowledge about baseball”
  + attempts to answer objective questions about baseball
    - "Which player on Red Sox contributed most to team’s offense?”, “How many HR's will Ken Griffey hit next year?”
  + cannot deal w/ subjective judgments (also important to the game):
    - “Who is your favorite player?” or “That was a great game.”
* **Statistics =**the best objective record of the game available
  + large part of Sabermetrics involves understanding \*\*\***how to use statistics\*\*\*** + which statistics are useful for what purposes, etc.
  + do NOT need to know a lot about math to understand sabermetrics, only some \*\*\***idea of how statistics can be used and misused\*\*\***
* Since Sabermetrics = an ***objective*** study of the game, it is necessary to \*\*\***use logical reasoning in arguments\*\*\***.
* **Hypothesis** can be developed from info you have, either from statistics or observations;
  + a claim which cannot be *directly* tested can be evaluated by studying conclusions which would follow.
  + \*Ex:\* “Pitching is X% of baseball,” which has been said with X between 15-80%.
    - want to test the claim “Pitching is 75% of baseball.”
    - If true, we'd conclude teams w/ best pitching = much more likely to win the pennant than teams w/ best hitting.
    - **However, this isn’t the case**.
    - League leaders in fewest RA (which is both pitching *and* fielding) win pennant about 1/2 the time + league leaders in RS (includes all of hitting) win just as often.
      * NOTE the definition of **offense** here = if you measure hitting by an **incomplete measure such as BA,**you'd conclude that pitching is much more important
    - Other unreasonable conclusions: team w/ 75% of its value in pitching would never trade a regular pitcher for a regular hitter, thus the claim must be rejected.
    - But if 75% is replaced by a number close to 40%, conclusions become reasonable.

# General Principles

* goal of a baseball team = win more games than any other team.
* 1 team has **very little control over # of games other teams win**, so **goal = win as many as possible**
* Therefore, it is of interest to measure player contribution to the team wins
* Clear relationship between team's RS and RA allowed + its W/L  (not perfect, but very strong.)
* Good formula, determined empirically from data by Bill James, is a team’s W/L ratio = square of the ratio between RS and RA
  + Thus a team which scores + allows the same # of runs will win + lose the same # of games, finishing at .500
  + team which scores 800 runs + allows 700 wins 64 games for every 49 it loses, which projects to a 92-70 record over a season (comes very close to actual records of most teams)
* **Basic goal of sabermetrics = evaluate a measure for a given purpose.**
  + **\*\*\*most common uses of statistics\*\*\* = evaluate past performance** (determine who should win MVP ) + **to predict future performance** (evaluate a trade that was just made).
  + In both cases, interested in **measuring contribution to games won and lost**.
* Baseball statistics can measure individual performance, ***independent of what other players do***
  + While importance of an individual event depends on situation, the effect of the situations on the importance of the statistic over a large sample such as a season is not great.
    - When a batter hits a single, this describes what *he* did; when a quarterback throws a ten-yard pass, the guard who took out a linebacker gets no statistical credit.
    - batter who received a single is properly credited for a success; the 10-yd. pass may have been a failure if it was 3rd down w/ 13 yards to go
    - Thus it is reasonable for the **goal of a baseball statistic to be to measure a player’s individual contribution to runs or wins.**
* Given goal, it's possible to **evaluate a statistic**.
* \*\*\*Baseball statistics can be evaluated, have same types of flaws, + be misused/misinterpreted in the same ways as *non*-baseball statistics\*\*\*
* 1st natural question to ask about a statistic = "*Does the statistic measure an important contribution to that goal*?”
  + \*Ex:\* ERA = # of runs a pitcher allows == *almost all a pitcher contributes to winning games*
  + BA does fairly well b/c it counts hits, but it ignores power + walks, which are also important parts of offense.
  + Few statistics fail badly here:
    - those which measure things happen rarely (like HBP)
    - those having little to do w/ winning games (ex: % of a batter’s outs that are SO's)
    - or both
    - Non-baseball \*Ex\*: # of crimes in a *city* last year = important if you want to know something about safety of city; # of crimes on a *single street* says very little about safety of whole city.
* 2nd + usually most important, question to ask is, “**How well does the statistic measure the player’s own contribution**?”
  + Many ways a statistic, baseball-related or not, can fail here.
  + \*\*\***Virtually every statistic fails in some way to some extent**\*\*\*, so the \*\*\***best statistics are those w/ only minor failings + relatively few of them\*\*\***
  + \*Ex:\* Player should be evaluated for what *HE*does, not for what teammates or manager do = a major problem w/ such statistics as RS
    - Unless a batter hits a HR or steals home, he needs teammate contributions to actually score, + cannot do much to cause them to get hits once on base.
    - Thus, batting in front of best HR hitters = score a lot, whether or not you have a good ability to score runs.
    - If batting you 8th on an NL team, won’t score many runs when you do get on base.
* **Good statistic should not measure *outside* effects over which player has no control (e.g. the park)**
  + Good non-baseball \*Ex:\* = high death rate in Miami.
    - population of Miami = older than population of most other cities;
    - thus, regardless of quality of medical care in Miami, we expect a high death rate.
  + Likewise: easier to score in Fenway than in Oakland
    - Therefore, pitcher w/ 3.60 ERA in Oakland could pitch just as well in Fenway, helping his team win games just as much, but have a 4.00 ERA.
  + Will sometimes see "park-adjusted numbers" = designed to eliminate this effect;
    - Ex: pitcher above might have a 3.80 park-adjusted ERA in either park.
      * Note: this is adjusting for value of the pitcher’s performance, not the actual performance;
      * 4.00 ERA for a Red Sox pitcher is just as valuable to his team regardless of how it is split between home + road games.
* **\*\*\*If a player’s statistics change considerably when changing teams, parks, or lineup positions, this suggests the outside effect has a major effect on the statistics.\*\*\***
* **\*\*\*If the statistic remains consistent when outside conditions change, this means it is measuring the player’s own contribution\*\*\***
  + Pitchers w/ good ERA’s tend to keep them when changing teams, so park effect is not a serious problem.
  + Hitters who score a lot in the leadoff spot score many fewer runs if dropped to 6th in the lineup, which means runs scored were mostly created by lineup position rather than the batter.
* **In addition to problems w/ outside effects, there can be problems with measurement.**
  + **\*\*\*No statistic can be useful w/out proper context/a measure of opportunities\*\*\*.**
  + There were more crimes committed in NY than in Boston last year, but this doesn’t say much about relative safety of the cities;
    - * to make such a comparison, you would need to compare crime ***rates***.
  + If a batter has 150 hits, ***what does that mean***? => If 500 AB's, he is good at getting hits; if 650 AB's, he is poor.
* **\*\*\*This is a problem w/ most counting statistics\*\*\***
  + BA places hits in a reasonable context, + this is recognized b/c batting title goes to player w/ the highest BA, NOT player w/ most hits.
* Similarly**, \*\*\*a statistic may not be useful if it tries to measure something with a very small sample size or number of occurrences\*\*\***
  + best pitchers at throwing shutouts often don’t lead league in shutouts, b/c league leader normally has ~5, + it’s quite common for a pitcher who usually throws 3 shutouts a year to get 7 in 1 year.
  + In contrast, best SO pitchers DO lead league in SO's (or SO's per 9 innings), b/c their totals are in the hundreds, + a pitcher capable of getting 250 strikeouts in 240 innings might get 230, but not 150.
* Same problem comes up w/ non-baseball statistics.
  + If 2/3 of people polled in a city plan to vote Democratic, it means nothing if it was 4/6 + not much if 40/60, but quite a lot if 400/600.
  + **\*\*\*This is the major flaw w/ many statistics often used on TV\*\*\***
    - statistic such as, “Wade Boggs is hitting .154 against Baltimore pitchers w/ runners in scoring position” means nothing b/c sample is probably 2 hits in 13 AB's
  + Sabermetricians agree w/ most fans that such stats are ridiculous + are there only to hold interest of (mostly statistically illiterate) TV audience.
* Once you have some idea of how well a statistic measures player’s own contribution to the goal, the \*\*\*final question to ask is, “**Is there a better way to measure the same thing**?”\*\*\*
  + *A statistic which has problems w/ the other questions but has no reasonable alternative measurement may still be useful.*
  + In contrast, a statistic such as RS, which can be *replaced by other statistics = very little value.*
  + A player’s own contribution to his total RS can be measured by his ability to get on base (already measured very well by OBP), +, to a lesser extent, to advance himself once he gets on base (measured by extra-base hits, + by stolen bases + caught stealing).
* Now, given these criteria, you can **evaluate a statistical conclusion**.
  + If you dispute the conclusion, your argument may be valid if based on these criteria
    - i.e. need to find something NOT measured by the statistic, or IS measured but *shouldn’t be*
    - \*Ex:\* can argue Mike Schmidt is a good hitter, even though his career average = .267, b/c he hit 548 HR + drew 1507 walks.
    - These are valid arguments, b/c **BA gives the same value to homers + singles**, + *does not count walks at all.*
    - Likewise, Ozzie Smith is *not a great offensive* player, but is *still an excellent* player, b/c of his *defense* (**no *offensive* statistic measures his *overall* value)**
* But you *CANNOT* dispute a statistical conclusion w/ a claim based on something that is *already included in the statistic*, or something which is *improperly measured* by your claim.
  + *NOT* reasonable to say Brooks Robinson was great at getting hits b/c of his 2848 hits;
  + *Correct* measure of how well he got hits = his .267 BA, which led to such a high hit total b/c his other skills allowed him to have a very long career
* **Turning 1 of the above examples around, you CAN'T claim Schmidt could not possibly be a great hitter, despite his .527 SLG, by looking at his BA, as BA is already counted in the slugging average.**

# Sabermetric stats

* A good, *complete* measure of individual offense would satisfy criteria above for a valuable statistic better than any of the traditional offensive measures.
* Therefore, sabermetricians often use/develop such statistics.
  + For measuring pitching: less need for such a statistic, b/c ERA + RA already count # of RA *by a pitcher*
* At team level, a **good measure of offense should have a strong correlation with RS**
  + i.e. should be possible to predict RS reasonably well from such a measure;
    - best teams by this measure should score a lot, while worst teams should score little
  + Measures such as BA do NOT do this
    - common for teams w/ best BA to be below average in RS.
  + RS *itself* obviously measures *team* offense very well, but creates a problem when trying to measure *individual* contributions;
  + it isn’t easy to measure directly how much a batter helped/hurt his team score.
* Several ways to develop a statistic which measures team offense.
  + Probably most natural way = say " score by getting runners on base *+ then advancing them"*
  + **Thus, a team’s RS should be *proportional* to # of runners it gets on base AND to the frequency w/ which it advances the runners.**
* OBP measures # of runners on base, while SLG (slugging average) =1 way to measure advancement
  + Note: an out reduces SLG, b/c it makes it less likely that any runners on base will be advanced
* Thus team **RS should be correlated w/ OBP\*SLG.**
* The \*\*\***test** of a statistic of this type **= how well it** **agrees with reality**\*\*\*
  + If you compare teams OBP\*SLG to their RS, you find a very good correlation w/ a standard error of just 24 runs.
  + For comparison, SD of RS in 1 season = ~70 runs (error returned if you predicted that all teams would be average in RS)
  + Meanwhile, BA alone has a standard error of 54 runs.
  + **The 24-run Std. Error covers everything that OBP\*SLG does NOT measure or measures improperly;**
    - includes such factors as baserunning + imperfections in the formula, but **much of the difference is chance.**
* Now, need to make an individual statistic by measuring a player’s contribution;
* OBP\*SLG is NOT the correct measure for a player b/c *he usually doesn’t drive himself in.*
* **Instead, want to multiply *his* OBP by the *team’s* SLG, + *his* SLG by *team’s* OBP.**
  + Since league (+ individual teams’) SLG are usually ~1.2 times OBP, **each point of a player’s OBP has 1.2 times the effect on OBP\*SLG that a point of his SLG has**.
  + Thus our measure **= (1.2\*OBP)+SLG.**
  + For simplicity, we often ignore factor of 1.2 + refer to **OPS, On-base Plus Slugging**
* When using **OPS**, remember that **OBP is slightly *undervalued*** + that SB's have NOT been counted.
* Using same process for other models of offense gives other measures, which give slightly different values for different elements of offense.
* The **choice of which measure** to use **depends** on which ones you **have handy**, the **purpose** for which you want **to use** it, + some **personal preferences**.
* But if you use any well-designed measure of offense, you won’t be wrong.
  + May find that a player w/ 2 more **Runs Created** than another is 003 worse in OPS, but such differences aren’t important
  + either way, you will reach the reasonable conclusion that they are very close.
* The *complete* measures of offense give a good estimate of the *value of the individual categories,* such as walks, HR’s, + outs, which make them up.
* The value of a player’s HR’s = the effect they have on OPS, or any similar statistic, + the importance of HR’s thus depends on this value + their frequency.

# Evaluating official statistics

* We can now apply the criteria to the official statistics
* Not reasonable to go through the arguments for *every* statistic, but useful to look @ statistics which cause the most frequent arguments.
* **RBI =** commonly used as a measure of a player’s offense, b/c they’re the only statistics which are **easily available** which look **like a complete measure**
  + *As a result, MVP is more likely to be league leader in RBI than in any other category*
* Of course, they are NOT a complete measure 🡺 ability to drive in runs is an *important* part of offense, but **not the whole thing**.
  + **This does NOT make RBI’s meaningless, only incomplete**.
* Real problem w/ RBI’s = the 2nd question (**How well does the statistic measure the player’s own contribution)**
  + RBI’s measure a lot of things which are NOT the player’s *own* contribution.
  + CANNOT drive in runners who are *not on base* (except w/ HR’s), but *your own batting doesn’t put them there*;
    - bat behind good players = get a lot of chances.
  + In fact, league leaders in RBI = much more likely to be players **who batted w/ the most teammates on base or in scoring position** (*not the batter’s contribution*) + NOT those who hit best w/ runners on base or in scoring position.
  + **Thus RBI = a better measure of who had the most chances to drive in runners than of who was the best at driving in runners.**
* 3rd test = **Is there a better way to measure the same thing**
  + There IS a better measure of ability to drive in runners
  + **Hits** drive runners in from scoring position + therefore, a player w/ many hits = good at this part of driving runners in.
  + Likewise, **extra-base hits** drive runners in from 1st base, + HR’s drive in from home plate.
  + ***Slugging average (SLG) =*** *player’s ability to get hits, extra-base hits, + HR’s, so it measures his ability to drive in runs, w/ park effects = the only significant bias.*
  + **Thus, RBI’s are NOT as useful a measure of offense, or even a measure of ability to drive in runs.**
* Other statistic subject to many of the same problems = pitcher’s W-L record (compare it to ERA)
  + Both measure something which is *clearly important*, since a pitcher’s goal = win games
  + the way he does this = preventing opponents from scoring.
  + *But both have some problems measuring the pitcher’s own contribution*
  + a comparison of their value depends on these problems.
  + 1st problem = runs are allowed by the WHOLE defense, not JUST by the pitcher
    - slightly more of a problem with W-L, as ERA eliminates runs due to **errors**, but NOT due to fielders that’re out of position, run slowly, or make weak throws.
  + At the MLB-level it isn’t a serious problem
    - good pitchers can still have good ERA’s (+ RA) even w/ teams of poor fielders.
  + **W-L record is one of the few categories which is immune to park effects**
    - there is 1 win in every game in every park.
    - ERA has slight problem w/ park effects 🡺 makes it more useful w/ park adjustment
  + **Most important factor = effect of the team offense.**
    - **\*\*\*Offense has almost no effect on ERA, but has a considerable effect on W-L\*\*\***.
    - A game is not won just by the pitcher (despite name of thestatistic), but by the team which scores more than allows.
    - In a single season, pitcher w/ best W-L record in the league = just as likely to be pitcher w/ best run support as the pitcher w/ fewest RA.
    - The run support is NOT pitcher’s contribution (except for batting in NL).
    - If there were pitchers who could cause teammates to score more for them, it would make sense to give the pitchers some of the credit.
    - But there is no tendency for pitchers who had support better than their team’s average in 1 season to have it again in the following season.
    - Nor does a pitcher have any control over whether he gets to pitch on a good offensive team
  + B/c of effect of run support, single-season W-L records are NOT a good measure of pitcher’s own value.
  + **ERA is available + is a better measure of what you actually want to know**.
  + **However, a career W-L reduces luck in run support by using a much larger sample size**.
  + In addition, pitchers rarely spend full careers w/ poor or good teammates.
  + Thus, a career W-L for a long career (several hundred decisions) = decent measure of pitcher’s own performance
    - it’s about as useful as a career ERA without park adjustments.
* Have now dealt w/ most common measures of batting + pitching + makes sense to now deal w/ most common measure of fielding.
* **Fielding average** = problem w/ first test (*Does the statistic measure an important contribution to that goal)*
  + While *defense is important*, an *incomplete measure of defense is NOT*.
* League leader in errors @ third usually makes about 30; leader in fielding average makes about 10.
  + These aren’t enough plays to make a difference of very many runs.
* More important part of fielding = **ability to prevent hits;**
* If 3rd baseman can’t reach a ball in the hole, or knocks it down but has no play, he won’t be charged w/ an error, *but the batter will get a hit which has the SAME effect.*
* **Errors are about as useful as a measure of defense as SO’s are as a measure of batting average.** 
  + They measure ONE way to fail to make a play;
  + while it is the most obvious failure, **all failures count the same on the scoreboard**.
* Fielder w/ poor range = a poor fielder whether he makes few or many errors, just as a hitter who hits too many grounders or popups can be a poor hitter even though he puts ball in play.
* While fielding average also has problems w/ park effects + scorer’s biases, the \*\*\***incompleteness is the most serious problem**\*\*\*
* Still, since it DOES measure SOMETHING useful, + fielders who are good at other things tend not to make errors (fielding % has a good correlation w/ wins), it would be a useful measure **in the absence of anything else**.
* So, it still has *some* value, particularly in concluding that players w/ very low fielding averages can’t handle their positions, but *it should be used in conjunction with putouts, assists, and an attempt to understand any biases in the numbers.*
* For recent players, better measure of overall defense = **Defensive Average (DA) =** makes fielding average unnecessary.
  + Basis for DA = division of playing field into zones of responsibility for fielders.
  + When a ball is hit into a fielder’s zone, it is **charged as an opportunity for that fielder**
    - if fielder turns it into an out, he receives credit for a play made.
      * Thus, all ground balls near 3rd are charged as chances for 3rd baseman;
        + a good 3rd baseman will make plays on most of them.
    - If he *fails* to make a play, effect is the same whether his throw is wild (error) or late (scored a single)
  + **fielding average does not tell you anything more.**
  + **Defensive average should be put to the same tests as any other statistic**
    - does reasonably well in the 1st test 🡺 measures a player’s ability to turn balls in play into outs, which covers *most* of his defensive play but *not all*
      * such skills as turning a double play + throwing out runners trying to stretch hits are NOT counted
    - does well in 2nd test (although still has some problems, mostly w/ park effects)
      * Pitchers cannot introduce bias simply by being left-handed (thus allowing a lot of ground balls to 3rd base + fly balls to left)
      * Good pitchers *may* help fielders’ DA *slightly* by allowing fewer hard-hit balls
      * Fielders do NOT have a great effect *on each other’s DA*, although there will be a *small* effect for plays such as low throws a good 1st baseman can handle
      * **All these effects will cause problems w/ almost any measure of fielding**
    - For 3rd test, DA = best measure of ability to make a play in the field *that we have*
      * isn’t perfect, but is *complete enough + accurate enough* to be useful.
* Thus the established statistics, used for reasons of tradition, may be *good* measures (ERA) or poor measures (RBI’s)
* \*\*\*Their value does NOT depend on *tradition* or *names*, but **depends on how well they meet the basic tests of any statistic**\*\*\*

# Other sabermetric arguments

* **Similar analysis must also be used in evaluating a hypothesis which depends on a statistical argument**
* **If a hypothesis leads to conclusions which don’t correspond w/ real game of baseball, it needs to be revised.**
  + Ex: natural question in predicting a player’s future performance in the major league = how useful minor-league #’s will be in a prediction.
  + problems w/ using minor-league #’s b/c there are *extreme* park effects + differences between leagues.
  + However, once you \*\*\*adjust a player’s minor-league #’s for these effects, + then make a specific adjustment for the difference between AA/AAA ball vs. majors, you *may* have something meaningful\*\*\*
  + There IS a method for making these corrections = **Minor-League Equivalency (MLE)**
    - will be useful if it works when tested against real world.
    - In fact, works almost as well as past major-league performance in predicting future major-league performance.
    - Most players w/ MLE’s which say they will hit .300 will hit close to .300 as rookies, just as most players who hit .300 last year will.
      * **Of course, neither prediction is perfect**
* Another issue sabermetricians have studied + discussed = **existence of clutch hitters**.
  + Clutch hits themselves certainly exit, but many players have reputations as players who will hit best w/ the game on the line
  + this is a hypothesis which CAN be tested; *are there any players with such an ability?*
  + Again, it’s necessary to **look at what actually happens, + what would happen if there were no clutch ability at all or if clutch hitting was a significant ability.**
  + Even if a .250 hitter were just a pair of coinswhich got a hit when both were heads, some .250 hitters would hit.400 during 1 season in the late innings of close games (3% chancein 80 AB)
  + So, the existence of such #’s doesn’t prove anything.
  + Butif there IS an ability, players who hit well in clutch in the past will continue to do so.
  + This CAN be tested, + HAS been == only*very weak evidence*, + but is clear that whatever abilitythere IS *does not mean much in baseball terms*.
  + There may be .267hitters who are actually as valuable as .268 hitters b/c of good clutch #’s
  + **But if replacing .268 w/ .275 = a conclusion inconsistent w/ what actually happens**

# VI. Conclusion

* Baseball statistics = useful **only if they enhance understanding of the game**.
* Therefore, they **should be judged by how well they measure what actually happens in the game**.
* **Meaningless statistics should be ignored or replaced + deficient statistics improved**.
* **Well-designed statistics should be used as an important part of discussion about the game + players**