

Statistical Baseball Research Bibliography
Charlie Pavitt

The goal of this essay is to introduce the Statistical Baseball Research Bibliography and explain its use. The Bibliography is the result of a comprehensive survey of published baseball literature. Along with many books, it includes articles in baseball journals (most notably the *Baseball Analyst*, *Baseball Research Journal*, and *By the Numbers*), academic sports journals (*Journal of Sports Economics* and *Journal of Quantitative Analysis in Sports*), and other academic journals (for example, *Operations Research*) and a few websites.

Articles have been included in the Bibliography if they meet the following criteria:

1 - They have been intended to make a contribution to our knowledge about baseball as a statistical science. This does not mean that the article must include statistical analysis. Many worthy articles have made theoretical or critical contributions without performing statistical analyses.

2 - While articles that present methods for evaluating or ranking teams or players are included if they make a meaningful contribution, articles that do nothing more than evaluate or rate teams or players are not included. Books that appear to have been intended to do little more than exploit the popular market for books on baseball statistics and make no contributions to our knowledge have not been included and will not be added in the future.

3 - Articles must either have been published in conventional markets (either academic or trade) or by SABR. Self-published works will only be included if they have made an unusual contribution (such as Cook's *Percentage Baseball and the Computer* and the early James *Baseball Abstracts*). The only exceptions are the *Baseball Analyst* and *By the Numbers*, as these represent the efforts of the statistical research community at large, and a few trustworthy websites.

The entries are arranged alphabetically according to the last name of the author of the article or book. Each entry has eleven columns of information. These are as follows:

Column A - Last name of first author.

Column B - First name and middle initial of first author.

Columns C and D present a code system identifying the content of the article. If an article includes more than one clearly different content area, it will be entered under each of them. In the case of books, it will be entered under each content area covered. The code system consists of two hierarchically organized levels, respectively called the macrocode, and microcode. Each of these codes is symbolized by one or two capital letters. To begin, each article is categorized within a general subject area. This general subject area is indicated by the article's macrocode, which can be found in Column C. Each general subject area is divided into more specific content areas. Each specific content area is indicated by the article's microcode, which can be found in Column D. It is important to remember that the same microcode may symbolize a

different category for different general areas. For example, the microcode S indicates Stolen base within the macrocode category Strategy, Starter/reliever within the macrocode category Pitching, and Succession within the macrocode category Managing.

Column E - The title of the article or book. Due to space limitations, the title may be shortened or paraphrased. If the title is not clearly indicative of the article's content and space permits, an indication of the content may be substituted, within { } brackets.

Column F- If journal article, title of journal, If book, name of publisher, If contributed chapter in book edited or mostly written by someone else, editor/author listed here. Look for listing of book under editor or primary author for title/publisher/year.

Column G - If journal article, volume or issue journal. If book, location of publisher.

Column H- Date of publication.

Column I- Pages that article is on. If article/book includes more than one subject area, pages will be specific to the discussion of the subject area if it is clearly distinct from the others.

Column J- Comments. If article is multi-authored, coauthors will generally be listed here. If article is part of debate, extension of earlier article, etc., other article(s) in series will be cited. When Column E consists of book title, title of book chapter may be listed here.

I update the Statistical Baseball Research Bibliography periodically with both new and old material that I find. Therefore, I would be interested in seeing any statistically-based articles anyone is familiar with not in the present version of the Bibliography and considering them for inclusion in future versions if they meet the criteria listed above.

The following presents the macrocode and microcode system. The code is indicated by the capitalized letter in each entry. Comments include the most popular areas of research under each category. This system was revised and expanded fairly recently, and I have not as yet completed a recategorization of books and baseball journals. For that reason, microcodes are currently missing for some entries, and some there now will be changed as I get to them.

Supercode	Macrocode	Comments
Batting Evaluation	true Ability	All nonsituational aspects of offense
	performance Consistency	Attempts to estimate
	run Expectancy	Measurement
	replacement Level	Type of evaluation method
	comparison among Methods	Type of evaluation method
	Other	Miscellaneous evaluation methods
	Predicting Performance	
	Ranking	Methods for measuring who is best
	Sample Size	And estimating performance
	Top-down	Type of evaluation method
	bottom-Up	Type of evaluation method
Batting Issues	Age/experience	Changes in performance over career
	Clutch	Does it exist? If so, how to measure it
	performance enhancement	Drugs Impact
	reaching on Errors	Differences among players
	Great feat odds	DiMaggio streak and others
	Historical changes	
	Interdependence	among players in batting
	inJury	its effects
	K - strikeouts	do they hurt more than other outs?
	bLack/white/Latin	Differences in performance
	Minor/major	Relation between the two in performance
	piNchhitting	As compared to being in lineup
	Protection	Does it exist?
	Replacement level	What is it?
hot/cold Streaks	Do they exist?	
Transactions.	Effect of changing team on performance	
Walks	Impact on offense; measuring batting eye	
Hit bY pitch	Who gets hit, who does the hitting	
Fielding	Catching	Methods for evaluating
	Double play	Impact on overall defense
	Evaluation	Methods for measuring how good
	Infielding	
	bLack/white/Latin	Differences in performance
	lOngevity	Differences across career
	Ranking	Methods for measuring who is best
	errorS	
outfielder Throwing	Methods for evaluating	
General	Introductions to baseball research. No macro or microcodes.	

Managing	Evaluation Succession	Methods for measuring how good Effects of changing managers on team
Overall player	All-star Evaluation Hall of fame Most valuable player Other Ranking	Total performance Predicting who gets chosen Methods for measuring how good Predicting who gets in Predicting who wins Miscellaneous issues Methods for measuring who is best
Pitching	Age/experience Clutch performance enhancing Drugs Evaluation Fly ball/ground ball height Historical changes Injury black/white/Latin Minor/major No hitters workload Power/finesse Ranking Starter/reliever Traded etc. Walks replacement level layoff	Its effects Does it exist? If so, how to measure it Impact Methods for measuring how good Methods for measuring Its effects Its effects Differences in performance Relation between the two in performance How likely? Who is likely to throw one? Analysis of ideal load Methods for measuring Methods for measuring who is best Comparisons; impact of relievers Effect of changing team on performance Their implications Measuring it Its effects
Situational	At bat Batter/pitcher matchups Day/night Fly ball/Ground ball Game Home/away Inning Left/right Opposition ball Park Runners in scoring position	Impact of ball/strike count Can one "own" the other? Impact on performance Impact on performance Runs scored by winners versus losers Impact on performance Runs expected given base/out situations Implications of platoon differentials Implications of who team is playing Impact on performance Impact on performance

	Season Turf/grass	Tendencies from month to month Impact on performance
STrategy	during At bats Batting order Fielding position sacrifice Hits Infield In pinch running Pinchhitting base Running Stolen bases platooning intentional Walks	Pitch selection Impact on performance Where should they play? When do they help? When does it help? Does it help? What type of player is best? Issues other than stealing Impact on offense; measuring ability When does it help? When do they help?
Team	Age/experience Competitive balance Disability Evaluation Hot/cold streaks black/white/Latin Pay and performance Ranking world Series Talent turnover Winning	Effects on team performance How often? How good? How quickly did team integrate? Standings What makes winners? Evaluation of how teams develop players How much? What is its impact? Factors involved
Umpire	Balls and strikes Ethnic bias Home plate	Calling them Does it exist? Variation among ump