

2014年度 S-PLUS & Visual R Platform 学生研究奨励賞応募

データで野球は語れるか？

-MLBデータからみた打撃の因子構造の解明-

専修大学人間科学部心理学科

心理統計学研究室

北條 大樹

八田 大輝

小川 泰史

目次

- I, はじめに
- II, 研究背景
- III, 研究目的
- IV, 研究方法
- V, 研究結果
- VI, 新たな疑問
- VII, まとめ
- VIII, 今後の課題・展望

はじめに

・そもそもMLBとは？

MAJOR LEAGUE BASEBALLという
アメリカの野球リーグの略称である



出典元:WALLPAPERS WIDE



出典元:billsportsmaps.com

はじめに

- どんな人がいるの？

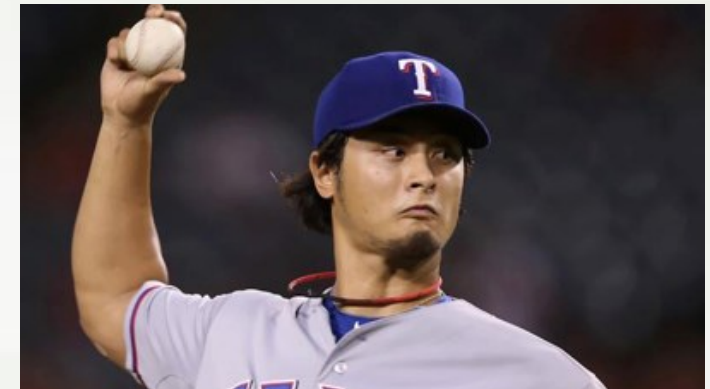
日本プロ野球からもイチロー選手や
田中将大投手、ダルビッシュ有投手を筆頭に活躍中



出典元:THE GOSPEL HERALD



出典元:bleacher report



出典元:MESN.com

- 今回の研究はLahmanパッケージ(Friendly et al, 2014)からデータを得た(Friendly et al, 2014, <http://cran.r-project.org/web/packages/Lahman/index.html>)

研究背景

セイバーメトリクス

セイバーメトリクス(Grabiner, 1994)

- 野球についての統計的研究
- チーム運営に役立てる球団も多い研究方法
- 勝利への貢献を前提に新たな評価指標を作成する
- 単なる初歩的な変数変換

因子分析

因子分析 (Spearman, 1904)

- 潜在変数が観測された変数に影響を与えるというモデルのもとで、変数間の関係を説明する分析
- 野球にも応用可能
- 現代統計学的議論や、予測などが可能になる

研究背景

野球データ

従来の方法

セイバーメトリクス
野球の統計的研究の筆頭候補
実践的な評価指標

今回の方法

因子分析
多変量解析の一つ
野球データへの実践例は少ない

研究目的

- 打撃成績から因子分析を用いて潜在変数を明らかにする
 - どれくらいの因子数なのか
 - どのような解釈になるのか
- セイバーメトリクスだけでは考えられなかったチームごとの特徴を推察する
 - 経年的にチームの変化を捉えられるのか
 - チームの変化を視覚的に分かりやすくできるか

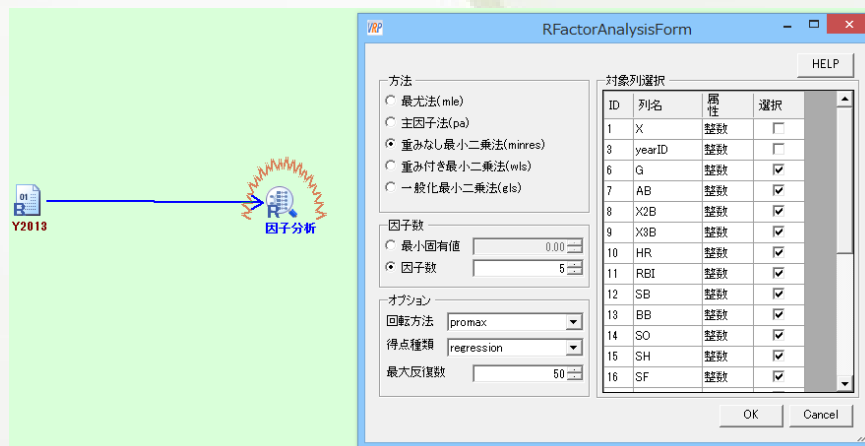
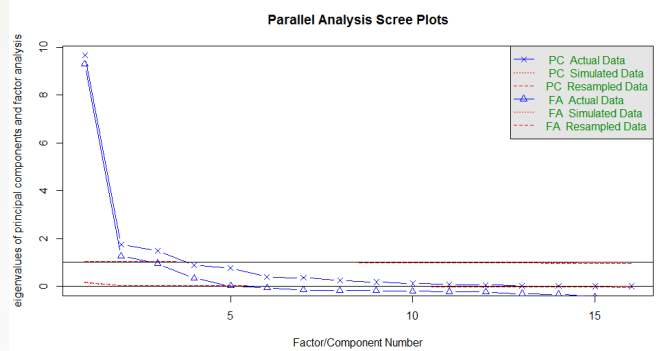


研究結果

分析方法

手法

- 実行には Visual R Platform (以下VRP) を使用
- 因子数の抽出方法は重みなし最小二乗法
- 平行分析を基に因子数は5に設定した
- 回転方法はプロマックス回転を使用



VRP実行画面

データ

- Lahmanパッケージと、それを用いて出塁率や打率等を表す変数を作成
- 2013年の打撃成績データを使用

分析結果

表1: 寄与率・累積寄与率

	MR1	MR2	MR3	MR5	MR4
SS loadings	8.340	2.260	1.731	1.484	0.741
Proportion Var	0.491	0.133	0.102	0.087	0.044
Cumulative Var	0.491	0.624	0.725	0.813	0.856

- 累積寄与率は5因子で85.6%だった
- 打撃能力の85.6%はこの5因子で説明できる

研究結果

因子を解釈する

表2: 因子負荷量(絶対値0.2以上 色付け太字)

		MR1	MR2	MR3	MR4	MR5	独自性
G	出場試合数	0.930	-0.030	-0.111	-0.037	0.139	0.153
AB	打数	0.925	-0.007	0.017	-0.010	0.119	0.005
X1B	1ベースヒット	0.832	0.001	0.030	0.240	-0.040	0.027
X2B	2ベースヒット	0.942	-0.037	0.012	-0.023	0.081	0.073
X3B	3ベースヒット	0.296	0.461	0.020	0.131	-0.042	0.413
MS	盗塁企画	-0.004	0.984	-0.012	0.026	0.014	0.005
RBI	打点数	1.088	-0.025	0.037	-0.038	-0.191	0.034
BB	四球数	0.901	0.038	-0.065	0.178	-0.097	0.131
SO	三振数	0.937	0.055	0.048	0.025	-0.139	0.123
HR	ホームラン	1.156	0.077	0.050	-0.573	-0.027	0.005
SH	犠打	-0.134	0.110	0.067	0.448	-0.036	0.775
SF	犠飛	0.833	-0.120	-0.015	0.114	-0.013	0.327
AVG	打率	0.061	-0.016	1.007	0.027	-0.135	0.054
OBP	出塁率	-0.020	0.000	0.821	-0.005	0.251	0.042
BBK	ストライク管理能力(*1)	0.132	-0.030	0.093	0.122	0.641	0.327
IsoD	四死球でどれだけ出塁したか(*2)	-0.069	0.026	-0.073	-0.053	0.975	0.197

*については付録にて説明

因子負荷量の高かった変数を抜粋した結果です。

MR1...ボールを遠くまで打つ力

MR2...走塁力

MR3...ボールにバットを当てる力

MR4...後続のバッターへつなげる力

MR5...粘る力

新たな疑問

Q:2013年では解釈しやすい因子構造がみられたが、
他の年でも同様に解釈しやすい因子構造はみられるのか？



- 1999-2013年の各年で因子分析を行った(詳細は付録にて説明)
- **結果...**
 - 同様の変数で因子負荷量が高いことが分かった
本研究により得られた因子構造の
安定性、妥当性が示された

新たな疑問

Q:各選手の因子スコアから、
各チームの因子スコア平均を算出し、
可視化することでチームごと・年ごとの特徴を
浮き彫りにできないだろうか？

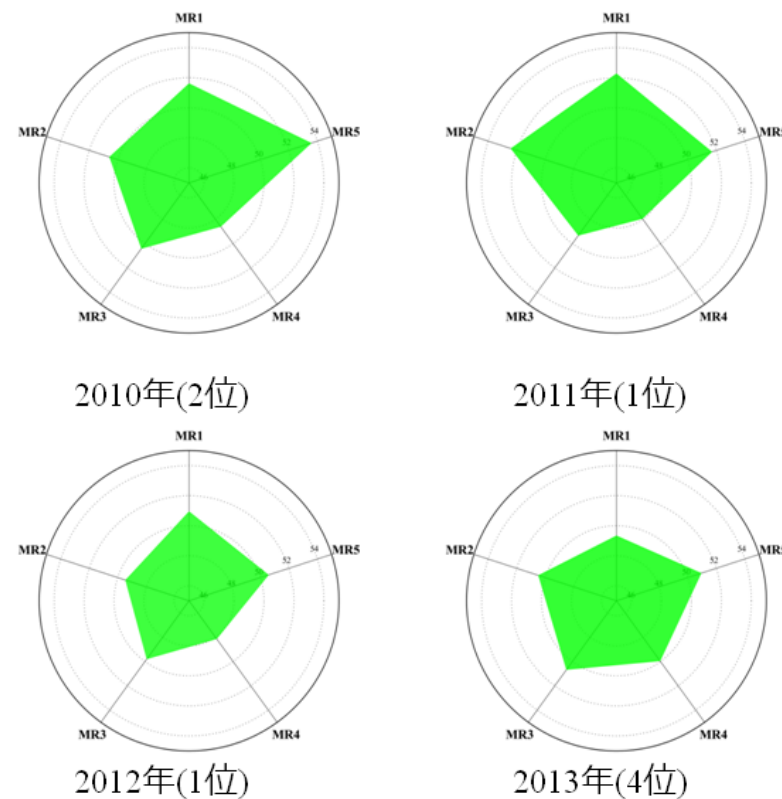
表3:各選手因子スコアの例

	playerID	teamID	MR1 SCORE	MR2 SCORE	MR3 SCORE	MR4 SCORE	MR5 SCORE
96601	aardsda01	NYN	-0.667	-0.395	-1.156	-0.886	-0.530
96602	abadfe01	WAS	-0.669	-0.395	-1.155	-0.885	-0.534
96603	abreuto01	SFN	-0.010	-0.165	0.867	0.004	0.426
96604	ackledu01	SEA	1.164	0.211	0.807	0.855	2.280
96605	adamsda02	NYA	-0.032	-0.394	0.386	0.279	0.287

- そこで各チーム各因子ごとの平均スコアを算出し、
レーダーチャートにした(代表的なチャートを次スライド)

ニューヨーク・ヤンキース(NYA)の 年毎レーダーチャート

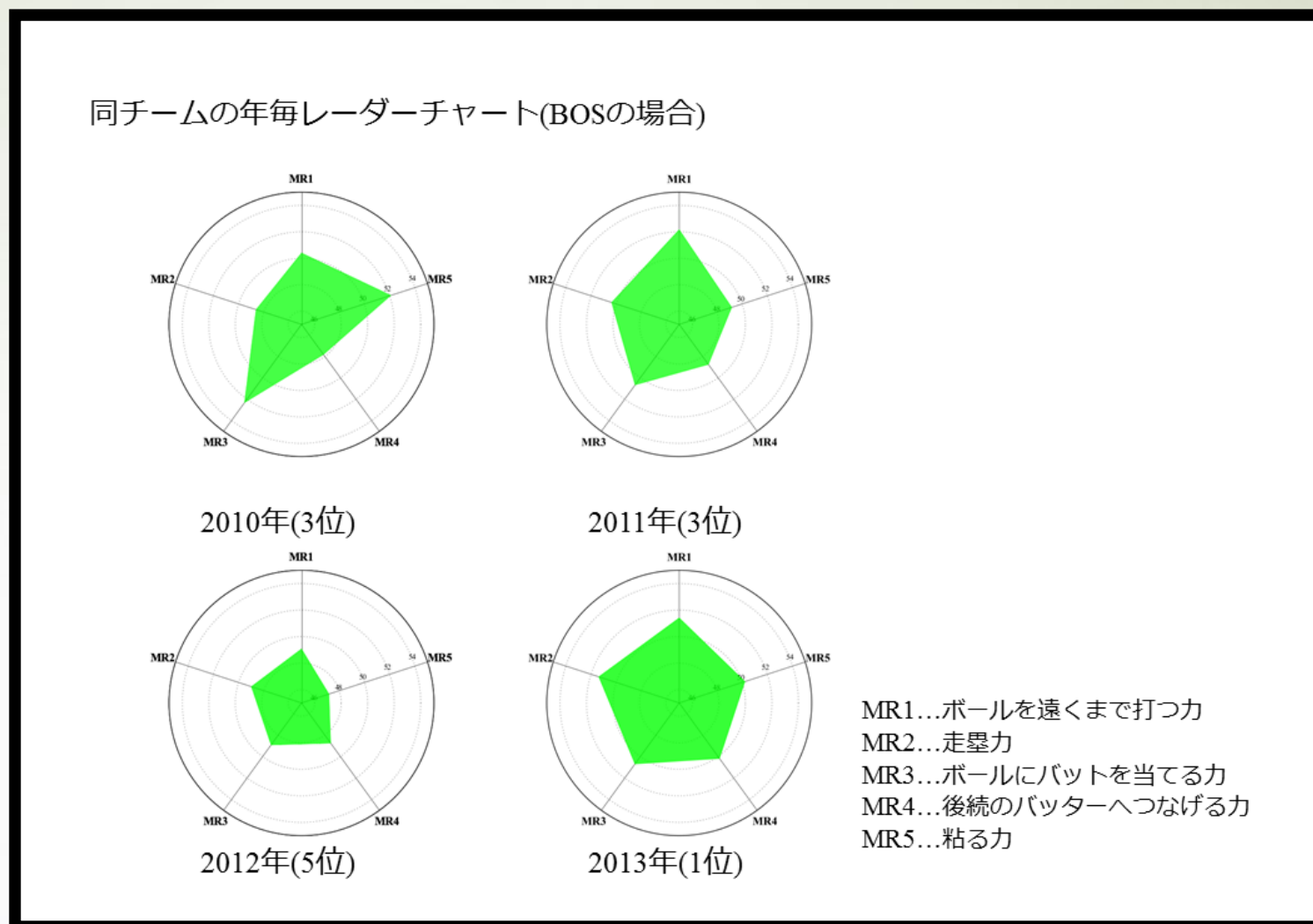
同チームの年毎レーダーチャート(NYAの場合)



VRPにより作成

NYAは
粘りながらコースの甘い球を強打していくスタイル('10~'11)から
単打を重ねていくスタイル('12~'13)へと変わっていった

ボストン・レッドソックス(BOS)の 年毎レーダーチャート



VRPIにより作成

BOSは

'11~'12にかけて大幅に因子得点が下がった結果、成績も下がった
'12~'13にかけて因子得点が上がった結果、成績も上がった

まとめ

- 打撃成績から因子分析を用いて潜在変数を明らかにした
 - 今回は打撃能力の潜在変数として、5つの因子が明らかになった
 - これらで85.6%の分散を説明することがわかった
- セイバーメトリクスだけでは考えられなかったチームごとの特徴を推察した
 - 因子得点のチームごとの経年的変化は経年的なチームの変化の特徴を推察できた

今後の課題

- 今回のデータには投手データを使っていない
 - 投手は打者よりも勝敗に影響するという俗説もあり、投手も分析することでチームの戦力を把握し、精度の高いリーグ順位の推測が出来るかもしれない
- 日本野球でも同様のことが言えるかわからない
 - 球場、ボールの大きさ、ストライクゾーン的位置等が違う日本独自の共通因子が存在するかもしれない

今後の展望

- ドラフトやトレードにおける判断材料
 - 各チームが分析することにより、わがチームは〇〇因子の選手が少ないため、もっとほしいなど。
- 日本国内のデータを使って、日本野球の順位予想
 - たとえば、今年は〇〇因子スコアの平均の高いチームAは優勝するであろう。
- 最大の目標...試合の勝敗予想
 - 今回のスタメンとベンチの平均因子スコアより、この試合はおそらく〇チームが勝つと推測される。さらに、走塁力因子が高いため、チャンスイニングでは大量得点が見込まれるといったことまで推測できる可能性がある。

謝辞

- まずは、このような学生研究奨励賞制度を設けて頂きましたこと並びに期間中にVisual R Platform の無償貸出しをしていただきましたこと、まことにありがとうございます。
- 私たちは、統計学を学んでいますが、Visual R Platform は、統計の知識が少しあれば、分析が非常に容易に行うことができ、同様にデータ処理やグラフ作成もGUI操作で行うことが出来ました。
- 実データ解析をするうえで大変素晴らしいソフトウェアだと実感しています。

付録

- 今回作成した変数

MS<-SB+CS #盗塁企画数
BBK<-BB/SO #ストライク管理能力(1つの三振で幾つの四球を取ったか?)
OBP<-(H+BB+HBP+IBB)/(AB+BB+HBP+IBB+SF) #出塁率
AVG<-H/AB #打率
IsoD<-OBP-AVG #四死球でどれだけ出塁したか

- 平行分析に使用したパッケージ

Psych: <http://cran.r-project.org/web/packages/psych/index.html>

付録

2000年から2013年までの因子分析結果(抜粋)

付録表1:寄与率・累積寄与率

	MR1	MR2	MR3	MR5	MR4
SS loadings	8.340	2.260	1.731	1.484	0.741
Proportion Var	0.491	0.133	0.102	0.087	0.044
Cumulative Var	0.491	0.624	0.725	0.813	0.856

付録表2:因子負荷量(絶対値0.2以上 色付け太字)

	MR1	MR2	MR3	MR4	MR5	独自性
G 出場試合数	0.930	-0.030	-0.111	-0.037	0.139	0.153
AB 打数	0.925	-0.007	0.017	-0.010	0.119	0.005
X1B 1ベースヒット	0.832	0.001	0.030	0.240	-0.040	0.027
X2B 2ベースヒット	0.942	-0.037	0.012	-0.023	0.081	0.073
X3B 3ベースヒット	0.296	0.461	0.020	0.131	-0.042	0.413
MS 盗塁企画	-0.004	0.984	-0.012	0.026	0.014	0.005
RBI 打点数	1.088	-0.025	0.037	-0.038	-0.191	0.034
BB 四球数	0.901	0.038	-0.065	0.178	-0.097	0.131
SO 三振数	0.937	0.055	0.048	0.025	-0.139	0.123
HR ホームラン	1.156	0.077	0.050	-0.573	-0.027	0.005
SH 犠打	-0.134	0.110	0.067	0.448	-0.036	0.775
SF 犠飛	0.833	-0.120	-0.015	0.114	-0.013	0.327
AVG 打率	0.061	-0.016	1.007	0.027	-0.135	0.054
OBP 出塁率	-0.020	0.000	0.821	-0.005	0.251	0.042
BBK ストライク管理能力(*1)	0.132	-0.030	0.093	0.122	0.641	0.327
IsoD 四死球でどれだけ出塁したか(*2)	-0.069	0.026	-0.073	-0.053	0.975	0.197

*については付録にて説明

因子負荷量の高かった変数を抜粋した結果です。

2013年

付録表3:寄与率・累積寄与率

	MR1	MR2	MR3	MR5	MR4
SS loadings	7.616	2.158	1.816	1.415	0.883
Proportion Var	0.448	0.127	0.107	0.083	0.052
Cumulative Var	0.448	0.575	0.682	0.765	0.817

付録表4:因子負荷量(絶対値0.2以上 色付け太字)

	MR1	MR2	MR3	MR5	MR4	独自性
G 出場試合数	0.822	-0.004	-0.082	-0.033	0.225	0.165
AB 打数	0.858	0.000	-0.001	-0.028	0.221	0.008
X2B 2ベースヒット	0.889	-0.043	0.008	-0.058	0.179	0.079
X3B 3ベースヒット	0.202	0.422	0.005	-0.036	0.260	0.409
HR ホームラン	1.143	0.066	0.046	0.004	-0.471	0.049
RBI 打点数	1.074	-0.028	0.032	-0.034	-0.123	0.021
SB 盗塁数	-0.013	1.013	0.006	0.013	-0.025	0.015
BB 四球数	0.863	0.051	-0.080	0.178	-0.006	0.115
SO 三振数	0.906	0.102	0.022	0.066	-0.112	0.107
SH 犠打数	-0.215	0.094	0.039	-0.010	0.471	0.796
SF 犠飛数	0.843	-0.068	0.039	-0.036	0.014	0.330
MS 盗塁企画数	0.035	0.956	0.000	0.010	0.020	0.013
X1B 1ベースヒット	0.752	-0.015	0.008	-0.088	0.386	0.021
BBK ストライク管理能力	0.131	-0.082	0.119	0.530	0.240	0.355
IsoD 四死球でどれだけ出塁したか	-0.053	0.026	-0.069	1.005	-0.059	0.144
AVG 打率	0.035	0.001	1.036	-0.150	0.023	0.030
OBP 出塁率	-0.002	0.012	0.838	0.222	-0.006	0.021

2012年

付録

- 2000年から2013年までの因子分析結果(抜粋)

付録表5:寄与率・累積寄与率

	MR1	MR2	MR3	MR5	MR4
SS loadings	7.392	2.202	1.530	1.307	0.788
Proportion Var	0.435	0.130	0.090	0.077	0.046
Cumulative Var	0.435	0.564	0.654	0.731	0.778

付録表6:因子負荷量(絶対値0.2以上 色付け太字)

	MR1	MR2	MR3	MR5	MR4	独自性
G 出場試合数	0.802	-0.027	-0.047	0.031	0.256	0.121
AB 打数	0.850	0.008	0.001	-0.007	0.234	0.007
X2B 2ベースヒット	0.918	-0.026	0.036	-0.056	0.128	0.067
X3B 3ベースヒット	0.220	0.463	0.010	-0.024	0.169	0.457
HR ホームラン	1.089	0.068	0.010	0.010	-0.432	0.056
RBI 打点数	1.051	-0.010	0.048	-0.054	-0.111	0.023
SB 盗塁数	-0.036	1.009	0.013	-0.015	0.009	0.013
BB 四球数	0.818	0.070	-0.084	0.247	-0.042	0.106
SO 三振数	0.850	0.068	-0.082	0.149	-0.002	0.126
SH 犠打数	-0.143	0.074	-0.053	0.069	0.483	0.776
SF 犠飛数	0.911	-0.134	0.050	-0.131	0.097	0.272
MS 盗塁企画数	0.018	0.963	0.010	-0.018	0.034	0.012
X1B 1ベースヒット	0.756	0.000	0.055	-0.090	0.372	0.022
BBK ストライク管理能力	0.138	-0.061	0.114	0.626	0.191	0.268
IsoD 四死球でどれだけ出塁したか	-0.060	-0.011	0.005	0.873	-0.032	0.308
AVG 打率	0.046	0.011	0.896	-0.073	0.011	0.212
OBP 出塁率	-0.040	0.022	0.828	0.163	-0.044	0.190

2011年

付録表7:寄与率・累積寄与率

	MR1	MR2	MR3	MR5	MR4
SS loadings	7.851	2.115	1.691	1.435	0.755
Proportion Var	0.462	0.124	0.099	0.084	0.044
Cumulative Var	0.462	0.586	0.686	0.770	0.815

付録表8:因子負荷量(絶対値0.2以上 色付け太字)

	MR1	MR2	MR3	MR5	MR4	独自性
G 出場試合数	0.888	-0.032	-0.066	-0.044	0.186	0.158
AB 打数	0.894	-0.020	0.007	-0.001	0.203	0.007
X2B 2ベースヒット	0.959	-0.070	0.014	-0.014	0.093	0.070
X3B 3ベースヒット	0.347	0.379	0.028	-0.046	0.139	0.454
HR ホームラン	1.091	0.040	0.031	-0.031	-0.452	0.049
RBI 打点数	1.059	-0.022	0.023	-0.025	-0.133	0.022
SB 盗塁数	-0.042	1.007	-0.008	0.007	0.016	0.014
BB 四球数	0.823	0.082	-0.062	0.228	-0.064	0.103
SO 三振数	0.899	0.077	0.037	0.008	-0.068	0.122
SH 犠打数	-0.157	0.090	0.032	0.011	0.510	0.728
SF 犠飛数	0.856	-0.056	-0.002	-0.006	0.019	0.313
MS 盗塁企画数	0.011	0.959	-0.005	-0.003	0.046	0.012
X1B 1ベースヒット	0.798	-0.020	0.012	-0.027	0.347	0.019
BBK ストライク管理能力	0.141	-0.059	0.027	0.730	0.169	0.238
IsoD 四死球でどれだけ出塁したか	-0.076	0.035	-0.015	0.891	-0.084	0.298
AVG 打率	0.046	-0.016	0.973	-0.112	0.036	0.104
OBP 出塁率	-0.024	0.009	0.855	0.190	-0.013	0.084

2010年

付録

- 2000年から2013年までの因子分析結果(抜粋)

付録表9:寄与率・累積寄与率

	MR1	MR2	MR3	MR5	MR4
SS loadings	7.294	2.191	1.758	1.316	1.059
Proportion Var	0.429	0.129	0.103	0.077	0.062
Cumulative Var	0.429	0.558	0.661	0.739	0.801

付録表10:因子負荷量(絶対値0.2以上 色付け太字)

	MR1	MR2	MR3	MR5	MR4	独自性
G 出場試合数	0.793	-0.036	-0.046	-0.048	0.282	0.150
AB 打数	0.816	0.001	-0.003	-0.020	0.277	0.006
X2B 2ベースヒット	0.885	-0.034	0.001	-0.024	0.170	0.066
X3B 3ベースヒット	0.241	0.444	0.003	-0.058	0.256	0.363
HR ホームラン	1.147	0.039	0.041	-0.012	-0.432	0.049
RBI 打点数	1.056	-0.034	0.019	-0.026	-0.079	0.019
SB 盗塁数	-0.032	1.018	0.001	0.011	-0.018	0.016
BB 四球数	0.857	0.080	-0.045	0.172	-0.057	0.124
SO 三振数	0.904	0.043	0.038	0.018	-0.026	0.116
SH 犠打数	-0.169	0.027	0.041	-0.024	0.592	0.714
SF 犠飛数	0.802	-0.077	-0.009	-0.043	0.159	0.295
MS 盗塁企画数	-0.012	0.966	-0.001	0.017	0.041	0.014
X1B 1ベースヒット	0.691	0.011	-0.006	-0.038	0.422	0.017
BBK ストライク管理能力	0.139	-0.054	0.062	0.551	0.220	0.405
IsoD 四死球でどれだけ出塁したか	-0.057	0.025	-0.050	0.950	-0.079	0.227
AVG 打率	0.037	-0.005	1.013	-0.148	0.040	0.053
OBP 出塁率	-0.010	0.009	0.846	0.217	-0.003	0.040

2005年

付録表11:寄与率・累積寄与率

	MR1	MR2	MR3	MR5	MR4
SS loadings	7.514	2.397	1.950	1.613	0.815
Proportion Var	0.442	0.141	0.115	0.095	0.048
Cumulative Var	0.442	0.583	0.698	0.793	0.841

付録表12:因子負荷量(絶対値0.2以上 色付け太字)

	MR1	MR2	MR3	MR5	MR4	独自性
G 出場試合数	0.823	0.248	0.123	0.164	0.261	0.152
AB 打数	0.881	0.288	0.175	0.164	0.281	0.005
X2B 2ベースヒット	0.888	0.222	0.171	0.148	0.202	0.071
X3B 3ベースヒット	0.488	0.424	0.108	0.072	0.330	0.456
HR ホームラン	0.928	0.129	0.151	0.130	-0.235	0.028
RBI 打点数	0.948	0.171	0.171	0.145	0.002	0.022
SB 盗塁数	0.327	0.908	0.078	0.089	0.185	0.021
BB 四球数	0.835	0.271	0.148	0.282	0.056	0.124
SO 三振数	0.856	0.230	0.165	0.187	0.128	0.136
SH 犠打数	0.093	0.220	0.077	0.040	0.448	0.735
SF 犠飛数	0.795	0.187	0.140	0.127	0.112	0.286
MS 盗塁企画数	0.369	0.893	0.088	0.100	0.211	0.005
X1B 1ベースヒット	0.833	0.332	0.184	0.145	0.341	0.025
BBK ストライク管理能力	0.371	0.146	0.253	0.643	0.144	0.343
IsoD 四死球でどれだけ出塁したか	0.161	0.048	0.154	0.872	-0.003	0.187
AVG 打率	0.235	0.087	0.927	0.080	0.100	0.062
OBP 出塁率	0.223	0.084	0.861	0.379	0.069	0.053

2000年

付録

使用画像引用一覧

- p.3 画像左 WALLPAPERS WIDE(http://wallpaperswide.com/major_league_baseball_mlb-wallpapers.html)
- p.3 画像右 Billsportsmaps.com(<http://billsportsmaps.com/?p=1558>)
- p.4 画像左 THE GOSPEL HERALD(<http://www.gospelherald.com/>)
- p.4 画像中 bleacher report (<http://bleacherreport.com/articles/2043712-masahiro-tanaka-wont-be-solved-by-mlb-hitters-after-dominant-start>)
- p.4 画像右 MESN.com(<http://nesn.com/2014/04/yu-darvish-becomes-fastest-pitcher-to-reach-500-ks-in-mlb-history/>)