

SEM によるパークファクターの検証と応用

森田 隼司 早稲田大学 基幹理工学部

板橋 智也 早稲田大学 基幹理工学研究科

塚田 将太、太田 文也、伴地 芳啓、直 玄峻 早稲田大学 基幹理工学部

三家 礼子 早稲田大学 理工学研究所

〒169-8555 新宿区大久保 3-4-1 早稲田大学理工キャンパス

Tel 03-5286-2853

E-mail: shunji-mrt.8@suou.waseda.jp

1. はじめに

野球では、球場のサイズ、形が様々であり、各球場で両翼、中堅、中間、面積、フェンスの高さ、グラウンドの種類（天然芝、人工芝、土等）が明らかに違う。これらの要因を解消するためにホーム、ビジターでの試合数で帳消しになっているはずだが、チームのホームでの試合数とビジターとしての試合数が同じだけで、各球場で均等に試合を行っているわけではない。例えば、巨人は東京ドームで 63 試合、他の 5 種以上の球場でビジターとして合計 62 試合行っている。これでは、完全な平等とは言えない。下記、12 球団のホームの形状である。重ねると大きさ、形は違う。従って、球場によりホームラン、三塁打、二塁打、の出やすさを客観的に比較するために「パークファクター」

の考え方が登場した。パークファクター自体の計算方法は簡単で、ホームランの出やすさに関わるパークファクター（以下 HR_{PF} とする）は、

$$HR_{PF} = \frac{\text{ホーム試合HR数} + \text{ホーム試合被HR数} / \text{ホーム試合全打席}}{\text{ビジター試合HR数} + \text{ビジター試合被HR数} / \text{ビジター試合全打席}}$$

となる^[1]。

そこで今回、パークファクターを視覚化できる構造方程式モデリング（以下 $sem^{[1][2]}$ ）のパス図を用いて各球団の比較を行い、パークファクターに値すると考えたホームランの推定値と実際公開されているホームランパークファクターとの相関関係を見ることで推定値の有効性を検証する。

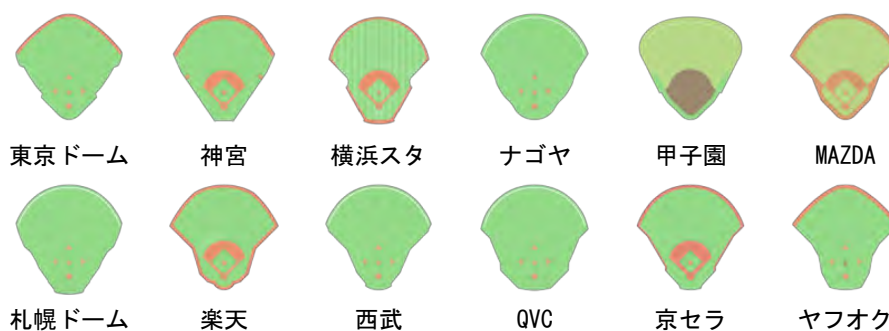


図 1 各球場の形^{[3][4]}

2. 使用データ

sem における、潜在変数と観測変数を決定するために貸与された野球プレイデータの変数からファイル選択を行う。その結果、使用可能な変数は打撃データにおける、「安打数」、「二塁打数」、「三塁打数」、「四球数」、「死球数」、「本塁打数」、「得点」

とした。前者 5 変数は潜在変数 1「打撃力」の観測変数、後者 2 変数は潜在変数 2「得点力」の観測変数とした。

3. sem の結果

各球団の sem のイメージ図を図 2 に示す。さらに、

図3に具体的な例として巨人のモデルを示す。

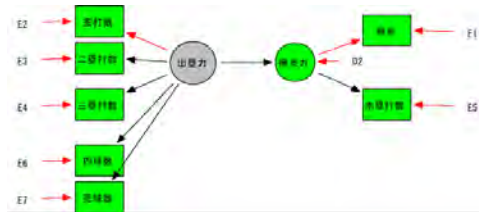
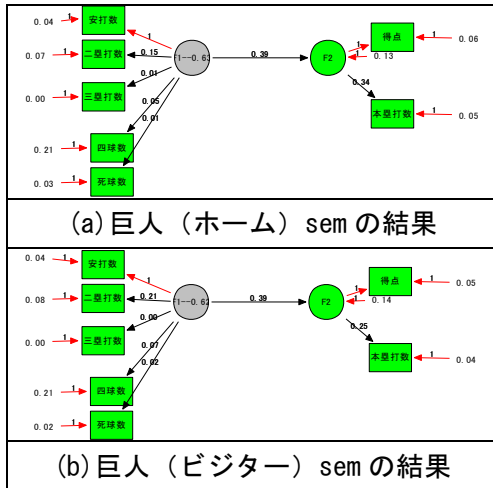


図2 SEMのイメージ図



(AGFI=0.9268,GFI=0.9647,RMR=0.0101)

図3 SEMの具体例

以上のように12球団すべてのホーム、ビジター別のパス図を描き、潜在変数から本塁打におけるパスの推定値を用いてsem独自のホームランパークファクタ(HR_PF_sem)を算出する。

$$HR_{PF_{sem}} = \frac{\text{ホームにおける本塁打へのパス推定値}}{\text{ビジターにおける本塁打へのパス推定値}}$$

12球団のHR_PF_semと2013年、2014年度に公開されているHR_PF^[3]との相関を見た。

(r=0.9167)

表1 PFの相関

HR PF ^[3]	HR_PF_sem
1.34	1.36
0.62	0.60
0.78	0.81
0.56	0.55
1.34	0.96
1.51	1.46
1.05	1.13
1.21	1.10
0.89	0.75
0.80	0.82
1.13	1.13
1.02	0.88

以上の結果より、従来のホームランパークファクターによりsemによるホームランパークファクターの有効性を検証した。

4. semからのHR(2B、3B)予測(単回帰モデル)結果
semの出力値より、各球団のホームとビジターにおけるホームランの予測を、単回帰モデルを用いて行う。算出の手順は以下である。

- ① 2015年のプロ野球データよりホーム、ビジターの各安打数を算出(プロ野球スルデータ置き場・2015年度版)
 - A) $a(\text{or } a') = \text{ホーム} \cdot \text{ビジター打率}$, $b = \text{全打数}$, $c = \frac{H(\text{or } V)}{H+V}$: ホーム・ビジター試合数、 $\text{安打数} = a(\text{or } a') \cdot b \cdot c$
 - B) $\text{安打数} = \text{全安打数} \cdot c$
- ② $F1 = \text{安打数} - E \text{安打数}$
- ③ 従属変数が潜在変数の構造方程式を導く…
 $F2 = p \cdot F1 + DF2$
- ④ 因子得点より潜在変数の単回帰モデルを作成…
 $F2 = \text{回帰係数} \cdot F1 + \text{定数}$
- ⑤ 従属変数が観測変数の測定方程式より、本塁打数を算出…
 $HR = q(\text{推定値}) \cdot F2 + E_{\text{本塁打数}}$ (HR分散共分散)

各安打数については(A)の計算方法を採用した。

以上より求めた、12球団の2015年度本塁打予測値と実測値を表2に示す。

表2 2015年度本塁打予測値と実測値

球団	予測値 (ホーム/ビジター)	実測値 (ホーム/ビジター)
巨人	74/50*	50/34*
阪神	66/65	29/40
広島	55/75	45/39
中日	36/62*	29/39*
DeNA	70/62	61/35
ヤクルト	74/59**	58/30**
ソフトバンク	41/52	67/51
オリックス	57/45**	50/34**
日ハム	58/56	39/55
ロッテ	42/47	30/42
西武	60/45*	68/48*
楽天	37/35*	34/34*

④における単回帰モデルの重相関係数は0.3~0.5なのでやや良いとする。

表の*は予測値と実測値のビジターに対するホームのHRの比率がほぼ等しい。**はそれに順ずる。12球団中6球団がパークファクターの影響が2015年も同様であると予測できる。

潜在変数の構造方程式の代わりに用いた単回帰(F1,F2)のグラフの例を示す。

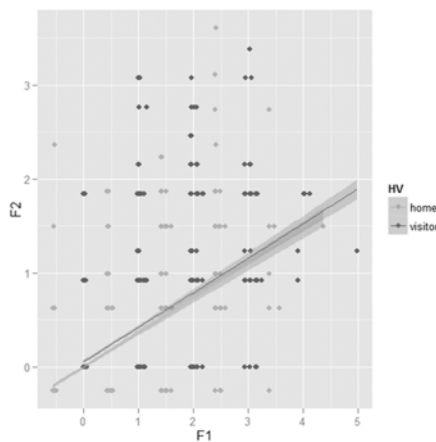


図3 潜在変数の単回帰曲線

ホーム :

```
lm(formula = F2 ~ F1, data = )
Residuals:
    Min       1Q   Median       3Q      Max
(1.4874) (0.0525) (0.0487) (0.0487) 2.7293
Coefficients:
(Intercept)  0.0000  0.0094  0.0010  0.9990
F1           0.3650  0.0115 31.8730 <2e-16 ***
---
Signif. codes:  0.0000 ***' 0.0010 **' 0.0100 *'

Residual standard error: 0.4282 on 2078.0000 degrees of freedom
Multiple R-squared: 0.3284, Adjusted R-squared: 0.3280
F-statistic: 1016.0000 on 1.0000 and 2078.0000 DF.
```

ビジター :

```
lm(formula = F1 ~ F2, data = )
Residuals:
    Min       1Q   Median       3Q      Max
(1.9423) (0.3324) (0.3324) 0.6442  3.6156
Coefficients:
(Intercept)  0.3324  0.0163 20.4100 <2e-16 ***
F2           0.8458  0.0274 30.8900 <2e-16 ***
---
Signif. codes:  0.0000 ***' 0.0010 **' 0.0100 *'

Residual standard error: 0.6754 on 2114.0000 degrees of freedom
Multiple R-squared: 0.311, Adjusted R-squared: 0.3107
F-statistic: 954.3000 on 1.0000 and 2114.0000 DF.
```

さらに二塁打(2B)、三塁打(3B)の予測を行う。算出の手順は以下である。(二塁打、三塁打ともに共通な式によって導き出される)

- ① 2015年のプロ野球データよりホーム、ビジターの各二塁打数、三塁打数を算出
本年度2015年のホーム、ビジター別の2B、3BがないためにHR予測で用いなかった安打の予測の①_B)をそのまま使用することとした。
 $2B = \text{全} 2B * c$
 $c = H(\text{or } V) / (H + V)$: ホーム・ビジター試合数
- ② $F1 = \text{安打数} - E \text{ 安打数}$
- ③ 2B (3B) の測定方程式より、
 $2B (3B) = q(\text{推定値}) * F1 + E_{\text{二塁打}} \text{ (二塁打分散共分散)}$

以上より、以下に12球団の2015年度2B予測値と3B予測値を表3に示す。

表3 2B、3Bの予測値

球団	2B 予測値 (ホーム/ビジター)	3B 予測値 (ホーム/ビジター)
巨人	89/112	4/2
阪神	125/97	11/18
広島	89/97	8/10
中日	127/103	11/12
DeNA	135/92	12/18
ヤクルト	114/109	7/12
ソフトバンク	114/120	19/17
オリックス	107/103	9/12
日ハム	118/117	11/15
ロッテ	113/136	21/13
西武	133/145	28/8
楽天	113/136	19/8

まず2B予測値と実測値について、2015年の各打撃に関わるPFは存在するので、簡易的に予測におけるホーム/ビジターの2B、3BによるPFを算出して比較する。実際の2015年PFを示す。

2015年セ・リーグの本拠地球場での各種PF

順位	チーム	スタジアム	得点PF	単打PF	二塁打PF	三塁打PF	本塁打PF
1	DeNA	横浜	1.10	0.914	1.14	0.451	1.24
2	巨人	東京ドーム	1.14	0.949	1.15	0.370	1.45
3	中日	ナゴヤドーム	0.984	1.09	0.870	1.57	0.903
4	阪神	甲子園	0.875	1.09	0.808	1.34	0.633
5	広島	マツダスタジアム	0.716	0.990	0.903	0.963	0.850
6	ヤクルト	神宮	1.22	0.978	1.39	0.412	1.58

2015年パ・リーグの本拠地球場での各種PF

順位	チーム	スタジアム	得点PF	単打PF	二塁打PF	三塁打PF	本塁打PF
1	西武	西武プリンス	1.01	0.939	0.912	0.367	1.04
2	日本ハム	札幌ドーム	0.939	0.873	1.27	0.820	0.788
3	ソフトバンク	ヤフオクドーム	1.33	1.09	1.18	0.745	2.30
4	ロッテ	QVCマリン	1.26	1.07	1.33	2.40	1.20
5	楽天	コボススタジアム	0.748	0.960	0.683	0.656	0.769
6	オリックス	京セラドーム	0.976	1.04	0.876	4.75	0.804

図4 2015年度12球団の各PF^[6]

次にHR、2B、3B予測値によるPF(sem_PF)

と各 PF の比較を示す。

これらの比較においては従来の PF の意味と同様で、PF>1 ならホーム優勢、PF<1 ならビジター優勢とする。この意味が、sem_PF と HR_PF の傾向が一致しているものを太字にしている。

表 4 2015 年度各 PF の比較

球団	HR PF		2B PF		3B PF	
	sem PF	HR PF	sem PF	2B PF	sem PF	3B PF
巨人	1.48	1.45	0.79	1.15	2.00	0.37
阪神	1.02	0.63	1.29	0.81	0.61	1.34
広島	0.73	0.85	0.92	0.90	0.80	0.96
中日	0.58	0.90	1.23	0.87	0.92	1.57
DeNA	1.13	1.24	1.47	1.14	0.67	0.45
ヤクルト	1.25	1.58	1.05	1.39	0.58	0.41
ソフトバンク	0.79	2.30	0.95	1.18	1.12	0.75
オリックス	1.27	0.80	1.04	0.88	0.75	4.75
日ハム	1.04	0.79	1.01	1.27	0.73	0.82
ロッテ	0.89	1.20	0.83	1.33	1.62	2.40
西武	1.33	1.04	0.92	0.91	3.50	0.37
楽天	1.06	0.77	0.83	0.68	2.38	0.66

HR については 12 球団中、広島・巨人・西武・中日・DeNA・ヤクルトの 6 球団で PF の傾向が一致した。

2B については、傾向が一致している球団は広島・日ハム・西武・楽天・DeNA・ヤクルトの 6 球団、となっていた。

最後に 3B については広島・DeNA・ヤクルト・日ハム・ロッテの 5 球団で傾向が一致した。ただし、3B についてはサンプル数の少なさが反映され、sem による PF でも実際の PF でも 2~3 の数字を示すものがあった。

5. sem からの HR(2B、3B)予測考察

- 全体的に長打の数は減っている
- 長打が出にくい要因が生じていると考えられる→ボールの反発係数など
- 3 塁打 PF は 3 塁打総数が少ないため、正しく評価出来ていないと考えられる。
- ソフトバンクの本塁打予測値がかなり外れていたが、やはり球場が小さくなった影響は大きい。
- ヤフオクドームの規格変更によりソフトバンクを除くパリーグ 5 球団の HR_PF は予測値よりも小さくなると考えられる。

6. まとめ

sem によるパークファクター（推定値）と実際のパークファクターとの間に強い相関があり、sem のパス図による可視化で、有効性の検証を行った。さらにホームラン、二塁打、三塁打の予測が sem により可能となった。

謝辞

本研究の貸与データは情報・システム研究機構の新領域融合研究プロジェクト『社会コミュニケーション』データ中心科学リサーチコモンズ事業『人間・社会データ』の支援を受けたものです。ここに感謝の意を表します。

参考文献

- [1] 山口・廣野(2011)「SEM 因果分析入門」日科技連
- [2] 田部井 (2011)「SPSS 完全活用法 共分散構造分析(Amos)によるアンケート処理」東京図書
- [3] 「古今東西の野球場についての各種データと野球場の大きさ、広さ比較」yakyujo.com 2015 年 12 月 1 日 閲覧
- [4] 「球場別広さ比較」
<http://www.d7.dion.ne.jp/~xmot/kyujohirosa.htm> 2015 年 11 月 1 日閲覧
- [5] 「球場のサイズや形は、記録にどんな影響を与えるか？」
<http://bb-nippon.com/column/1215-data/8638-20140729no5sabr?page=4> 2015 年 11 月 1 日閲覧
- [6] 「今年急激に本塁打が出やすくなったのは、やっぱりあのドーム！」
<http://www.baseball-lab.jp/column/entry/168/> 2015 年 12 月 1 日 閲覧

本著作物は原著作者の許可を得て、株式会社日本科学技術研修所（以下弊社）が掲載しています。本著作物の著作権については、制作した原著作者に帰属します。

原著作者および弊社の許可なく営利・非営利・イントラネットを問わず、本著作物の複製・転用・販売等を禁止します。

所属および役職等は、公開当時のものです。

■公開資料ページ

弊社ウェブページで各種資料をご覧ください <http://www.i-juse.co.jp/statistics/jirei/>

■お問い合わせ先

(株)日科技研 数理事業部 パッケージサポート係 <http://www.i-juse.co.jp/statistics/support/contact.html>