

NPBにおいてブレイクする選手はどのように予測できるのか？ ～1軍での活躍と2軍成績・選手情報の関係～

1. 研究動機

昨今、日本のプロ野球(NPB)では若手選手の活躍が大きな注目を集めている。活躍する選手の中には、1年目から1軍で活躍する場合もあるが、はじめのうちは2軍で経験を積みプロ入り後数年経ってから1軍の主力として活躍する場合も多い。また、2022年シーズンのオフには現役ドラフトが導入され、まだ1軍での活躍には乏しい選手の移籍がより活性化すると予想される。そのため、そのような選手の内、1軍で活躍する見込みのある選手(「ブレイク」選手)を予測することができればチームの効率的な戦力アップにつながるのではないかと考え、本研究を行った。

2. 本研究における「ブレイク」の定義

野手 「2軍での出場がほとんどであった野手が 翌年1軍で **多くの得点を生み出すこと**」

投手 「2軍での出場がほとんどであった投手が 翌年1軍で **多くの投球回投げること**」

3. 分析 → 重回帰分析 (ブレイクに対し、複数の要素が影響しているという仮説を立てたため)

野手分析 I

投手分析

目的: 野手のブレイクと2軍成績・選手情報の関係を明らかにする
方法: ①下記の目的変数・説明変数を標準化したもので重回帰分析を実行
②P値が最も大きい説明変数を除外し、再度重回帰分析
③①,②をP値0.05未満の説明変数だけによる回帰式が得られるまで繰り返す
対象: 2017~2021にウエスタンまたはイースタンリーグで**規定打席**に到達した野手100人
(2軍規定打席到達→その年は2軍での出場がほとんど→1軍では活躍できていない=ブレイク条件)

目的: 投手のブレイクと2軍成績・選手情報の関係を明らかにする
方法: ①下記の目的変数・説明変数を標準化したもので重回帰分析を実行
②P値が最も大きい説明変数を除外し、再度重回帰分析
③①,②をP値0.05未満の説明変数だけによる回帰式が得られるまで繰り返す
対象: 2017~2021にウエスタンまたはイースタンリーグで**試合数×0.6の投球回**に登板した投手83人(2軍でこの投球回→その年2軍登板ほとんど→1軍では活躍できていない=ブレイク条件)

目的変数: 野手ブレイクポイント(FBP) = $wOBA(\text{Speed}) \times \text{打席数}$
1打席に生み出した得点
説明変数: age(年齢), AVG(打率), IsoP(純粋な長打力), IsoD(選球眼), K%(三振率), SBP(盗塁ポイント)(=盗塁数×盗塁成功率) の6つ

目的変数: 投手ブレイクポイント(PBP) = 投球回
説明変数: age(年齢), throw(利き手), H/9(被安打率), BB/9(与四球率), K/9(奪三振率), LOB%(投手の粘り) の6つ

・収集データ:対象選手の2軍成績(2017~2021)とその翌年の1軍成績(2018~2022),および生年月日と利き手。
✳ ageとthrow→2軍開幕時点 FBP→1軍成績から その他の説明変数→2軍成績から
・通常1シーズン143試合制であるところが2020年のみ120試合制であったため、2020年のFBP,PBPへはさらに $\frac{120}{143}$ をかけたものを使用。

・収集データ:対象選手の2軍成績(2017~2021)とその翌年の1軍成績(2018~2022),および生年月日と利き手。
✳ ageとthrow→2軍開幕時点 PBP→1軍成績から その他の説明変数→2軍成績から
・throw…**ダミー変数** 右投=0,左投=1
・通常1シーズン143試合制であるところが2020年のみ120試合制であったため、2020年のPBPへはさらに $\frac{120}{143}$ をかけたものを使用。

<説明変数間の多重共線性のチェック>
age IsoP IsoD AVG K% SBP
1.2 1.4 1.2 1.2 1.2 1.1
図1:説明変数間のVIF

<説明変数間の多重共線性のチェック>
H/9 BB/9 K/9 LOB% age throw
1.3 1.3 1.3 1.1 1.3 1.0
図2:説明変数間のVIF

重回帰分析

重回帰分析

結果
F値: 0.0064 (≤ 0.05) P値: age→0.034 IsoP→0.0058 (≤ 0.05)
切片: 2.6×10^{-16} 偏回帰係数: age→-0.21 IsoP→0.28

結果
F値: 0.021 (≤ 0.05) P値: age→0.050 IsoP→0.050 (≤ 0.05)
切片: -6.5×10^{-17} 偏回帰係数: age→0.21 IsoP→0.21

検査

検査

独立性○ 等分散性○ 正規性✗
→この分析は適切ではない(残差の正規性が不成立!)
<残差のダービンワトソン検定> DW=1.9 $\neq 2.0$
<予測値と残差>
図3: 予測値と残差
図3より、残差は縦軸0を中心に均一に分散している。よって残差は**等分散性がある**といえる
<残差Q-Qプロット>
図4: 残差Q-Qプロット
図4より、点は曲線を描いた為、残差に**正規性がない**といえる

独立性○ 等分散性○ 正規性○
線形性○
<残差のダービンワトソン検定> DW=2.2 ≈ 2.0
<各説明変数の残差グラフ>
図5: 予測値と残差
図5より、残差は縦軸0を中心に均一に分散している。よって残差は**等分散性がある**といえる
<残差Q-Qプロット>
図6: 残差Q-Qプロット
図6より、点は直線を描いた為、**正規性がある**といえる

残差の正規性の改善

図8: FBP Q-Qプロット 目的変数を
図9: FBP' Q-Qプロット
自然対数変換
曲線 → 直線
非正規 → 正規

野手分析 II

残差が正規性をもたない際、目的変数に正規性をもたせると改善されることがある。
→図9よりFBPは非正規分布、図10より**自然対数変換**すると(FBP'とする)正規分布に!
→ **野手分析 I** の目的変数を FBP から FBP' に変更して再試行 (VIFは分析 I で確認済み)

結果
F値: 0.0029 (≤ 0.05) P値: age→0.048 IsoP→0.0018 (≤ 0.05)
切片: 4.4×10^{-16} 偏回帰係数: age→-0.19 IsoP→0.31
検査 独立性○ 等分散性○ 正規性○ 線形性○
<残差のダービンワトソン検定> DW=2.1 ≈ 2.0
<予測値と残差>
図10: 予測値と残差
図10より、残差は縦軸0を中心に均一に分散している。よって残差は**等分散性がある**といえる
<残差Q-Qプロット>
図11: 残差Q-Qプロット
図11より、点は直線を描いた為、残差に**正規性がある**といえる

<説明変数と目的変数の関係>
FBP' - age ... 相関係数-0.21(図15)
FBP' - IsoP ... 相関係数0.28(図16)
→ どちらも非常に弱い**線形関係がある**
図12: ageとFBP' 図13: IsoPとFBP'

<説明変数と目的変数の関係>
PBP - K/9 ... 相関係数0.22(図9)
PBP - throw ... 相関係数0.22(throwはダミー変数なので散布図省略)
→ どちらも非常に弱い**線形関係がある**
図7: K%とBP

4. 解釈と考察

野手

投手

野手のブレイク度を表す指標FBP'は下の式で予測できる。
 $FBP' = IsoP \times 0.31 - age \times 0.19 + 4.4 \times 10^{-16}$ (全変数標準化した状態)
野手のブレイクは影響度の大きい順に、2軍での**IsoP, age**と関係がある。(AVG, IsoD, K%, SBPとは無関係)
IsoPは大きい程、ageは小さい程、FBP'は大きく予測
『**長打力があり、若い選手**』ほどブレイクしやすい(長打力優先)

投手のブレイク度を表す指標PBPは下の式で予測できる。
 $PBP = K/9 \times 0.21 + throw \times 0.21 - 6.5 \times 10^{-17}$ (全変数標準化した状態)
投手のブレイクは、2軍での**K/9, throw**と同程度関係がある。(age, H/9, BB/9, LOB%とは無関係)
K/9, throwともに大きい程、PBPは大きく予測
『**奪三振能力があり、左投げの投手**』ほどブレイクしやすい
(非標準化式より、左投げのアドバンテージ=9イニング当たりの奪三振能力3.8個分)
→ 球団は『』内のような選手を優先して1軍に昇格・起用するべきである

5. モデル評価

図14: 野手分析 II の主なモデル評価指標
MAE 0.75 RMSE 0.94 RMSE/MAE 1.3
図15: 投手分析の主なモデル評価指標
MAE 0.72 RMSE 0.95 RMSE/MAE 1.3

MAE, RMSEはあくまで相対的なものだが、どちらの分析の数値からも改善の余地があると考えられる。具体的には、外れ値の影響を大きく受けて精度を落としている可能性などが予想される。一方、どちらの分析でもRMSEとMAEの比は $\frac{1.253}{1} (=1.253)$ に近い値を取っており、モデルがデータの特徴を十分表現していることが再確認できる。

6. 今後の展望

・ブレイク指標の改善 → 野手: 守備面も含めた指標(WAR等) 投手: 先発と中継ぎを平等に扱う指標
・モデルの精度向上 → 外れ値の除外 訓練データとテストデータに分割・評価
・予測の実践: 2022年2軍成績からの2023シーズンの予測
→ 野手: リチャード(ソフトバンク) 投手: 鈴木(ロッテ) が特に高く予測