BIとビッグデータⅠ　レポート課題　2112601 篠原　啓希

(1)

昨今日本プロ野球では、国内の独立リーグでプレーする選手が指名されることが年々増加している。独立リーグはトーナメントが主の社会人野球と違い、加盟チームが50~70試合程のリーグ戦を行い、独自に成績を付ける。また、NPB球団の2・3軍との交流戦をリーグ戦として年に10試合以上行うリーグもあることから、社会人野球よりも成績の信ぴょう性があり年間を通じNPB球団へのアピールができるなどのメリットがあり、実際に2023年のプロ野球ドラフト会議では育成契約を含め過去最多の23選手が指名されている。

しかし、社会人野球よりも十分な試合データがあるにも関わらずその成績からプロ野球でどのような成績を残すかの数値的な予測はあまり行われておらず、独立リーグの強みを活かしきれていないという問題があり、NPB球団側も入団後の戦力構想が狂うことや指導方針を変更することがあることが考えられる。

(2)

独立リーグは選手の入れ替えが激しく、複数年での予測は不向きだと考え、今回は昨年福岡ソフトバンクホークス3軍との交流戦を各チーム年12試合行い、データがまとまっている四国アイランドリーグの昨年の成績データを用いる。

昨年の四国アイランドリーグのチームの打者成績:https://data.iblj.co.jp/stats/2023/hitting

昨年の四国アイランドリーグのホークスの成績:https://data.iblj.co.jp/teams/23704/2023

昨年の福岡ソフトバンクホークスの2軍選手成績:

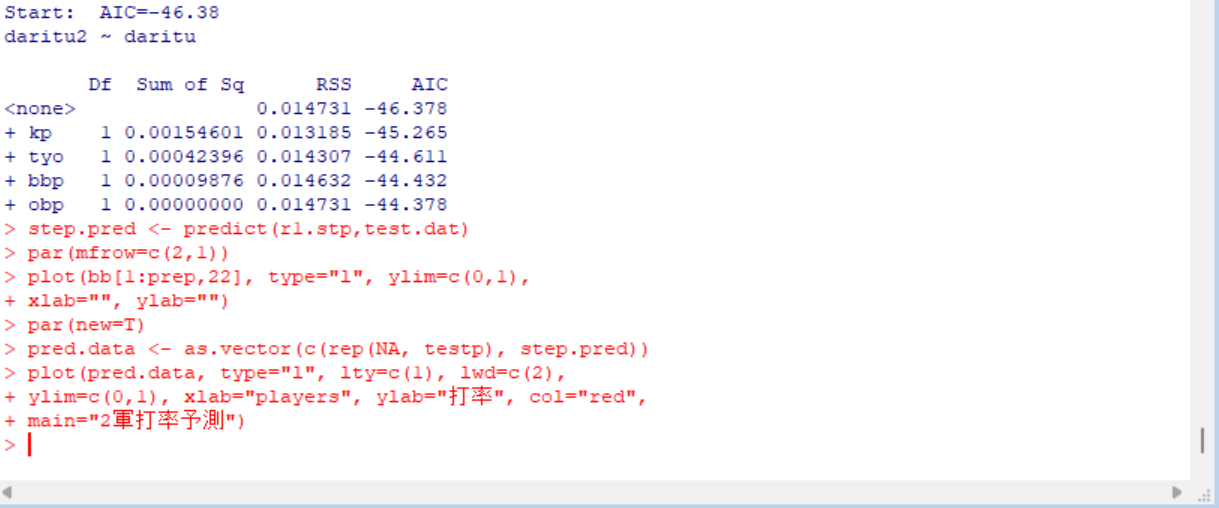
https://baseball-data.com/23/stats-farm/hitter2-h/tpa-1.html

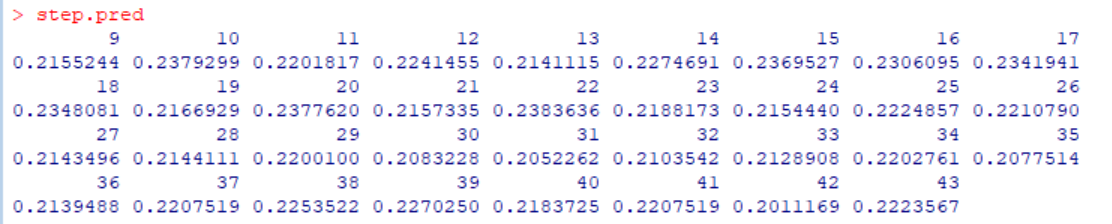
(3)

今回は打者成績の予測を行う。2軍との比較として、ホークス3軍の試合に出ている選手で2軍公式戦の成績の内、2・3軍合計で９０打席かつ四国アイランドリーグで35打席以上かつ2軍で45打席以上出場している8選手の成績をモデルの推定に、四国アイランドリーグのチームの選手の内、150打席以上出場した35選手のデータを用いて、選手の2軍成績を予測する。また、データとして打率、試合数、打率、打席、打数、得点、安打、二塁打、三塁打、本塁打、塁打、打点、三振、三振率、四球、四球率、死球、犠打、犠飛、併殺打、出塁率、長打率の内、安打数などは同じ率でも打席数で変わることを考慮し四国アイランドリーグの打率、三振率、四球率、出塁率、長打率を説明変数の候補として2軍での打率、三振率、四球率、出塁率、長打率をそれぞれ目的変数とし、ステップワイズによる重回帰モデルで予測する。

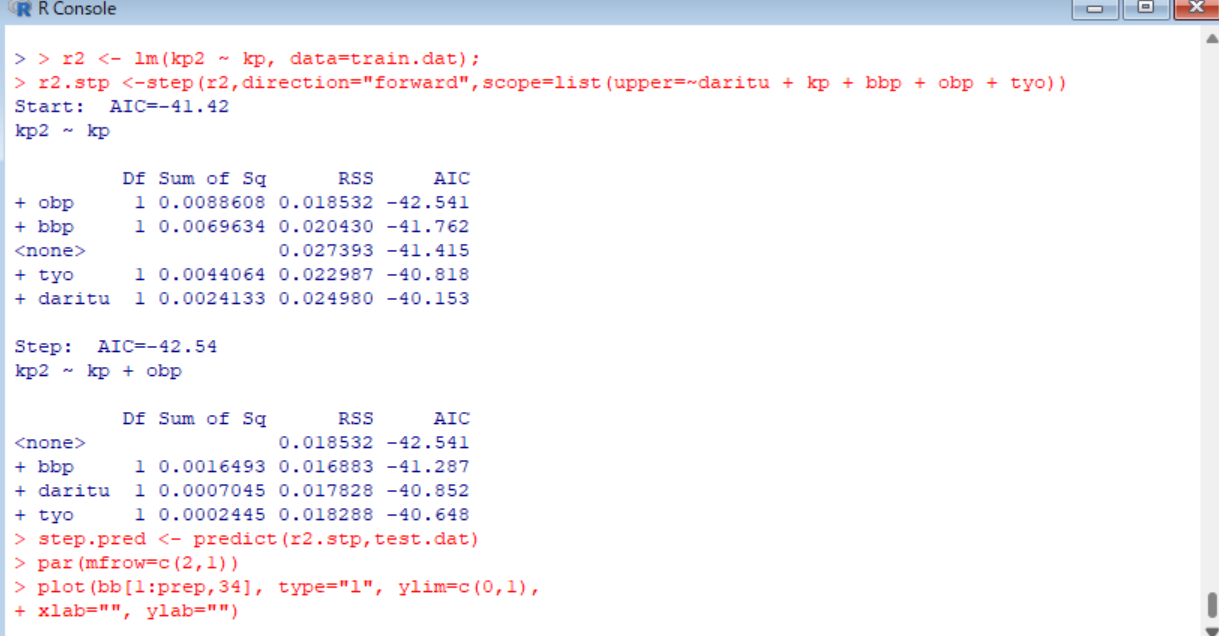
kpが三振率、bbpが四球率、obpが出塁率、tyoが長打率。

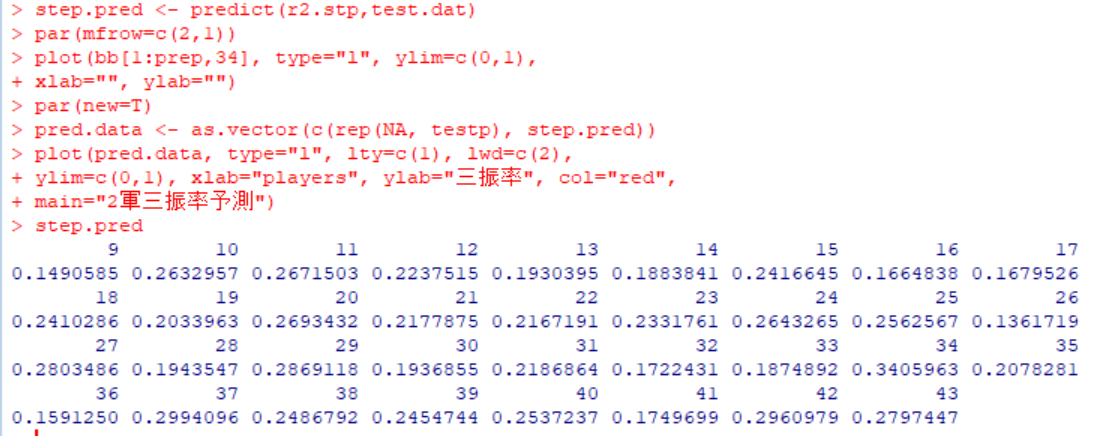




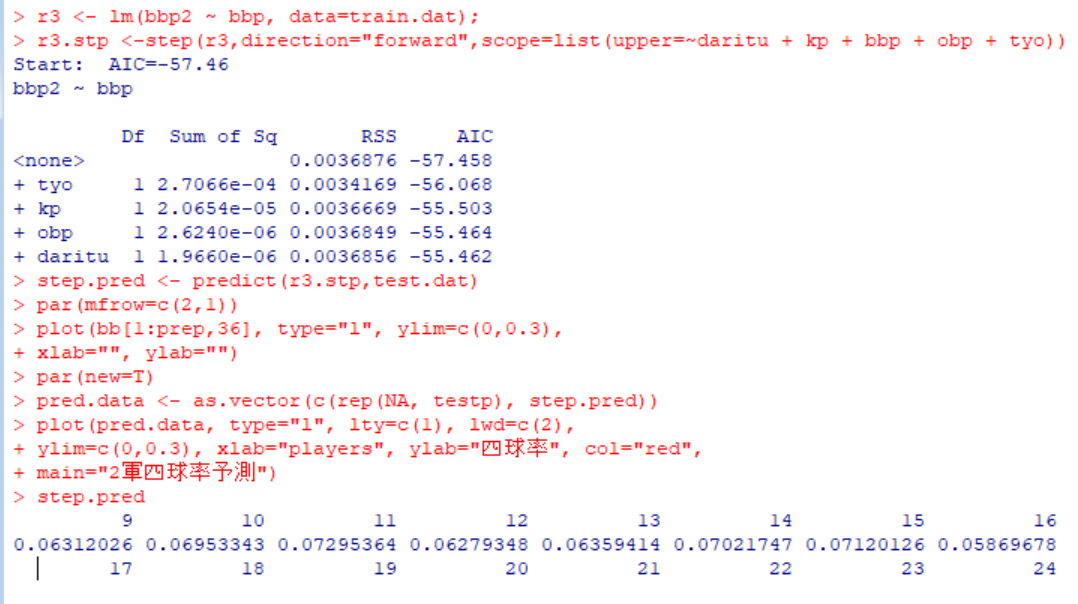


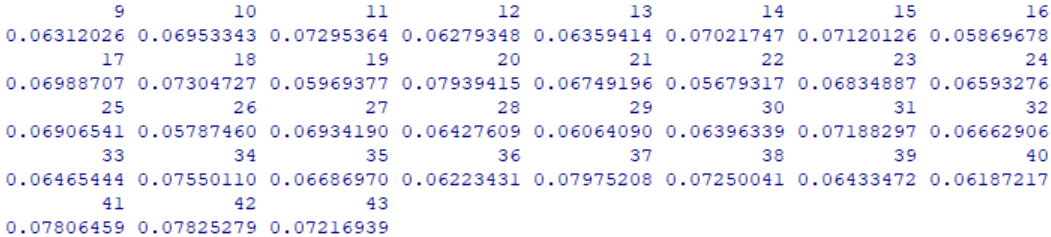




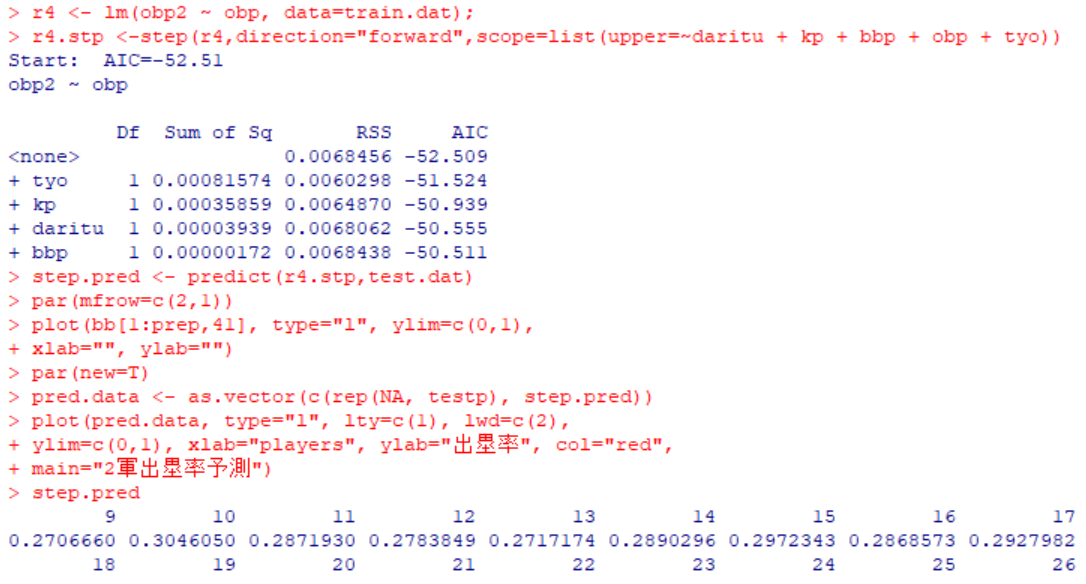


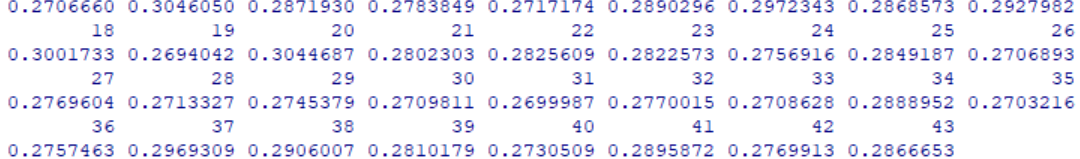






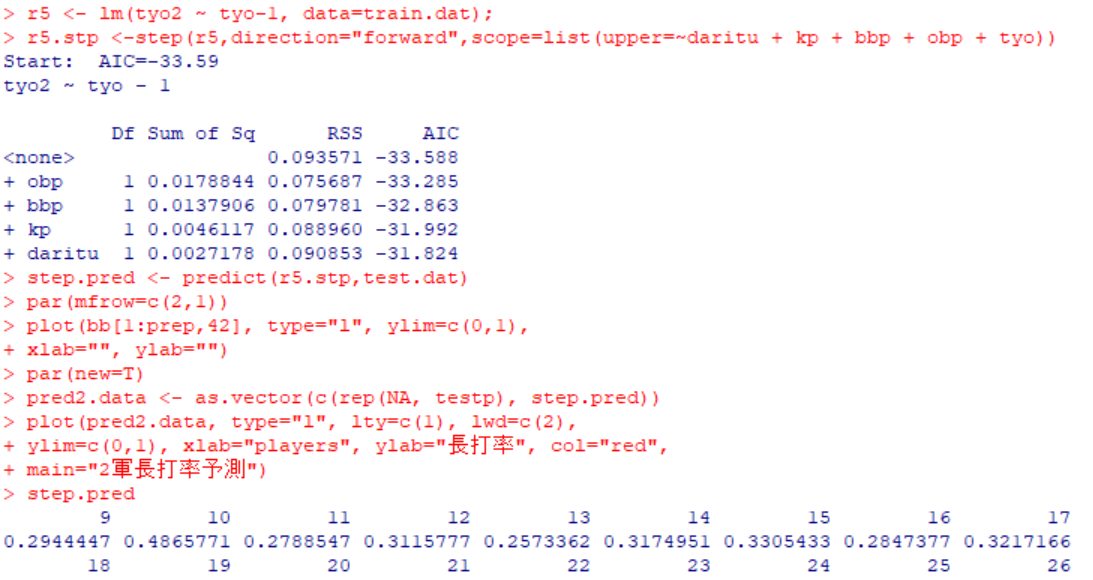


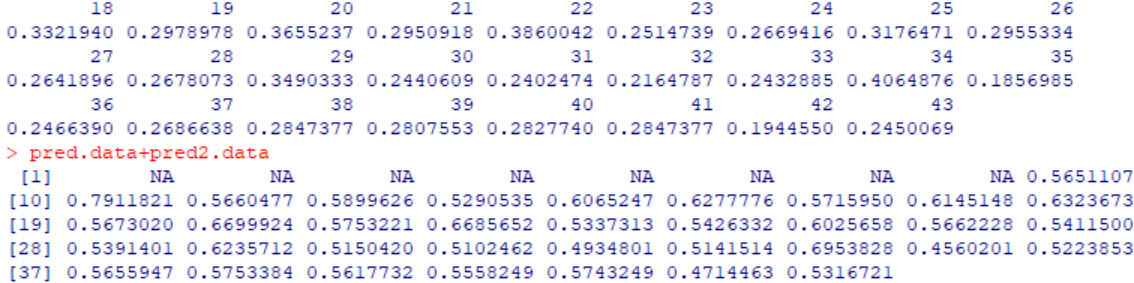






最後に一般に打者の貢献度を示す出塁率＋長打率=OPSを表示







(4)

昨年四国アイランドリーグからプロ入りまたは2軍のみの球団に入団し、2軍試合にすでに出場している選手の8/1時点の成績と比較し評価する。

井上絢登(DeNA)

予測(10人目)：打率0.238 三振率0.263 四球率0.0695 出塁率0.305 長打率0.430

今季：232打席打率0.291 三振率0.121 四球率0.0603 出塁率0.359 長打率0.456

増田将馬(くふうハヤテ)

予測(15人目)：打率0.237 三振率0.242 四球率0.0712 出塁率0.297 長打率0.338

今季：237打席打率0.274 三振率0.176 四球率0.0773 出塁率0.332 長打率0.335

河野聡太（オリックス）

予測(18人目)：打率0.235 三振率0.241 四球率0.073 出塁率0.300 長打率0.339

今季：88打席打率0.276 三振率0.102 四球率0.102 出塁率0.356 長打率0.316

打率に関しては予測が実際を0.04ほど下回り、三振率は1/2程になり、四球率は0.03~0.006ほど差があり、出塁率に関しては予測が実際を0.05ほど下回り、スケールを考慮するとかなり差があることからあまりいい予想ができていないといえる。長打率は誤差が0.03位内に収まっていて、比較的的確に予想ができているといえる。

今回は切片を考慮しないモデルの予想をしなかったが、考慮することで精度がよくなる可能性がある。また、今回は純粋な打力の評価を考えるため盗塁に関するパラメータを除いたが、走塁に関係ない三振率以外は考慮すると精度が上がると考えられる。

また、NPB球団1球団の選手だけではモデルの推定に使うデータがどうしても少なくなってしまうので、打席数を考慮せずモデルの推定に使うことで精度が上がるかなとも考えた。