

# 競艇選手の成績に関する統計的方法

2007MI061 市川範洋

指導教員：木村美善

## 1 はじめに

競艇とは、6艇のモーターボートを操る選手が、1周 = 600メートルの楕円の水上コースを3周し、着順を争う競技である。観客はレースの着順を予想し、勝舟投票券（舟券）を購入する。購入した舟券が的中した場合、的中券と引き換えに配当金を獲得する。どの選手が勝つかを事前に予測できれば競艇で負けることはないであろう。しかしながら、それはとても難しいことである。ならばどのような状況なら選手達が勝ちやすく、なおかつ強い選手というのは何が優れているかが分かれば、競艇というギャンブルに勝つ確率というのは上がるであろう。それを統計的に分析することで競艇をより詳細まで見られるようにするために本研究を行うことにした。

## 2 データについて

「ボートレースオフィシャルウェブサイト」[1]から、2010年11月から2011年4月までの36選手のデータを対象とした。データは変数を「選手」、「勝率」、「3連対率(%)」、「出走回数」、「優出回数」、「優勝回数」、「平均スタートタイミング」、「フライング回数」、「出遅れ回数」、「能力指数」、「コース別進入指数」、「級別」以上の12個を変数として用い主成分分析とクラスター分析を行った。（[4]参照）

## 3 分析方法について

本研究で用いた分析方法は、主成分分析、クラスター分析、数量化 類を用いた。

## 4 主成分分析

### 4.1 主成分分析とは

多くの変数の値をできるだけ情報の損出なしに、1個または少数個の主成分で総合力や特性を分析する方法である。多くの特性を、元の変数に含まれる情報を出来るだけ無くさず、多くの変数を少数個にまとめる合成変数を求めるための手段である。主成分分析の利用目的はこのように情報の縮約にある。

### 4.2 分析結果

第1主成分で累積寄与率は、75.3%、第2主成分で95%となり、80%を越えるので第2主成分までとする。

### 4.3 主成分分析の考察

主成分分析では、主にオッズが高いか低いかと、出走回数が多いか少ないかを分析することができた。出走回数が多い選手がよく勝つという傾向が見れたが、3連対率などからあまり出走しない選手でも勝率が高い選手が存在した。それは、オッズとしてはあまり高くはないかもしれないが、実力的にはある程度の力を持った選手と意味付けで

きるだろう。第2主成分では、負の方に突出した特徴の違う木村光宏(6)、広瀬政憲(27)の選手がいるが6は出走回数が最も多かった。27は出走回数が少なく3連対率が極端に低いためこのような結果になったと考えられる。

主成分得点のプロット図を以下に示す。

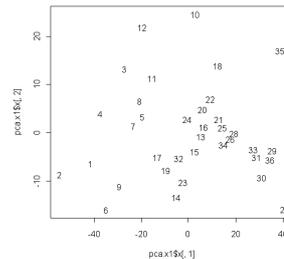


図1 主成分のグラフ

## 5 クラスター分析

### 5.1 クラスター分析とは

異なる性質のもの同士が混ざり合っている集団（対象）の中から、互いに似たものを集めて集落（クラスター）を作り、対象を分類しようという方法を総称したものであり、数値分類法とも言われている。（[1],[2]参照）

### 5.2 分析方法について

まず、それぞれがどのような選手なのか、特徴を調べるためにオリジナルデータをクラスター分析した。最も良い結果がでた最長距離方を用いて、2010年11月から2011年4月までの半年間の無作為に選んだ36選手を分析した。

### 5.3 オリジナルデータの分析結果

クラスター分析を行った結果、デンドログラム図の左から順に第1群から第4群に分かれた。（図2、参照）

### 5.4 オリジナルデータによるクラスター分析の考察

・第1群は、勝率・3連対率・優出回数が極めて低いことから、あまりいい順位に絡んでこないことが考えられる。

・第2群は、ややフライング回数が多いのだがその分平均スタートタイミングが他よりもいいことから、スタートダッシュの面で優れている。さらにコース別進入指数が他の群よりも高いことから、第2群は先行逃げ切りタイプの選手が多いと考えられる。

・第3群は勝率・優出回数・優勝回数が他の群よりもずば抜けていいことから、この選手は優勝戦線に残る群であると考えられる。

・第4群は、特に特徴的な部分がないことから標準的な選手の群だと考えられる。

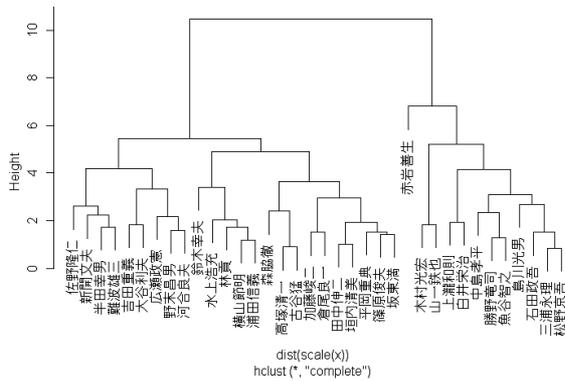


図 2 デンドログラム

## 6 数量化Ⅱ類

### 6.1 数量化Ⅱ類とは

数量化Ⅱ類とは、判別予測分析および要因分析のための質的データ解析の方法である。判別分析の定性データへの拡張ともいえるが、判別式から算出したサンプルスコアがどの群に属するか調べる。各要因について質的な測定量が得られる場合には、判別分析の方法が適用できる場合があり、その意味で質的（定性的）データによる判別分析といえることができる。カテゴリー質量は相関比（群間分散/全分散）を最大にするように定められる。しかし、説明アイテム間は独立関係をなるべく保つこと、外的基準とは、従属関係を保つことが重要である。（[3] 参照）

表 1 出力結果

アイテム		カテゴリー	範囲	偏相関係数
2 連対率	1	-0.1698	0.2902	0.2309
	2	-0.1296		
	3	-0.0359		
	4	0.0871		
	5	0.1204		
3 連対率	1	-0.6802	1.9874	0.9161
	2	-0.8329		
	3	-0.5286		
	4	1.1545		
	5	1.1212		
出走回数	1	0.4733	0.9110	0.7106
	2	0.1196		
	3	-0.4377		
	4	-0.2152		
	5	-0.3113		
優出回数	0	-0.2277	0.8398	0.5501
	1	0.0767		
	2	0.2284		
	3	0.2770		
	4	-0.3634		
	5	0.0558		
	6	0.0558		
	7	0.0476		
	8	0.1506		
	9	0.2453		
	12	0.2453		
	平均スタートタイミング	1		
2		0.1487		
3		-0.3045		
4		-0.1196		
フライング回数	1	0.0522	0.2091	0.3728
	2	-0.1568		
外的基準	50 %以上	1.2245		
	50 %未満	-0.7792		

## 6.2 考察

分析の結果から外的基準の「50 %以上」の方向を示したアイテムは、「2 連対率が 40 %以上」、「3 連対率が 60 %以上」、「出走回数が 60 ~ 100」、「平均スタートタイミングが 0.14 ~ 0.19」などの結果が得られた。範囲と偏相関係数の結果がどちらも 1 番高かったのは 3 連対率でその後に出走回数と続いた。これらのことから考えられることは、3 連対率の数値が他のアイテムと比べても圧倒的に大きいので平均的にどのレースでも 3 位以内に入っている選手が 1 位を取る確立というのが高いということが言えた。またカテゴリースコアのなかで平均スタートタイミングが 0.14 ~ 0.19 まだが 50 %以上の方向を示したことから、スタートで少しでも出遅れたタイムを出す 1 位はかなり取りにくいということも分析から見受けられた。逆に優出回数やフライング回数は、偏相関係数が高くなかったことから、1 位を取るための条件にはなりうるということも分析結果から言えた。他にもアイテムを状況や会場によって変えることで、どの条件が重なると勝ちやすいなどという結果も得ることができた。

## 7 まとめ

この研究を終えて、やはり 1 位を取る為には勝率はもちろんであるが、3 連対率や優出回数などからも優勝争いに絡んでくる選手が多いことが分かった。特に能力指数に関しては、主成分分析で第 1、第 2 両方共に関わってきていることから、重要性が伺えた。その後数量化Ⅱ類を行った結果として、1 位を取る為には何が重要であるかということが分かった。さらにアイテムを変えることでデータだけでは分からなかったことが見えてきてその日の状況や時間帯、天候などが変わることによって重要なアイテムが変わってくることも分析できた。この研究を通して競艇選手というのは勝つためにはある程度の経験を積むことで上位に入ってくる結果としてとても現れた。また競艇というモータースポーツは特にスタートからの第 1 コーナーまでのタイムが 1 位を取るための重要な要素であることも分析から見受けられた。このことから、中盤での逆転というのは難しいとも言える。なので競艇というモータースポーツは、ベテランに優位でかつスタートを得意としている選手が優勝するための条件を 1 番満たしているのだと思う。

## 参考文献

- [1] 青木繁伸：「R による統計解析」．オーム社.2009.
- [2] 金 明哲：「R によるデータサイエンス」．森北出版, 2007.
- [3] 内田治：「数量化理論とテキストマイニング」．日科技連出版社, 2010
- [4] BOAT RACE OFFICIAL WEB SITE:  
<http://www.boatrace.jp/>