

< 修 士 論 文 >

競馬において人気と結果に
乖離のあるレースの予測

滋 賀 大 学 大 学 院
デ ー タ サ イ エ ン ス 研 究 科
デ ー タ サ イ エ ン ス 専 攻

修了年度：2021年度

学籍番号：6020123

氏 名：山田 宏輝

指導教員：梅津 高朗

提出年月日：2022年1月12日

目次

第1章	はじめに.....	2
	研究の背景・目的.....	2
	分析使用するデータ.....	2
	分析の方法.....	2
第2章	中央競馬の仕組み.....	4
第3章	分析.....	6
第4章	結論・考察.....	8
	謝辞.....	9
	参考文献.....	9

第1章 はじめに

研究の背景・目的

近年、再度盛り上がりつつある競馬であるが、これの結果を予想することは困難である。下馬評道理決まるレース、そうでないレース様々あるが、人それぞれ好みのレースを探して的中を目指している。しかし、人気な馬が勝つと予想してそのまま人気の馬が勝つて的中したとしても、払戻金は穴馬が勝ったときと比べて小さくなる。そして、払戻金が投票した金額を下回ってしまうと回収率は上がらない。そこで、なるべく人気な馬だけが来るような堅いレースではなく穴馬も来るような荒れるレースを見つけようと考えた。また、競馬についての概要は第2章にて述べる。

分析に使用するデータ

日本中央競馬会が提供している JRA-VAN データラボにて公開されている「TARGET frontier JV」から取り込んだ公式のデータを使用する。また、JRA の東京競馬場の過去 15 年間で行われた 7477 レースを分析対象とする。障害レースはレース回数が少ないため今回は除外した。

分析の方法

荒れるレースを分析するために過去のレースで 1, 2 着に入った馬の平均人気がどのような要素で変化するかを明らかにするために重回帰分析を用いる。よって、本論文では 1, 2 着馬の平均人気を目的変数として、その他レース開始前の時点で正式に公開されている情報を説明変数とする。レースデータの一部は以下の通りとなっている。

表 過去 15 年間の東京競馬場でのレースデータの一部

平均人気	距離	良	芝	頭数
-0.63797	-0.84985	1	0	-0.7137
2.298296	1.442136	1	0	0.490365
-1.12735	-0.195	1	0	0.490365
1.31954	-0.84985	1	0	-1.11505
-0.39328	1.114709	1	1	-2.72047
0.830162	-0.195	1	0	-1.91776
-0.63797	-0.84985	1	0	0.490365
-1.12735	-0.84985	1	1	-2.31911
-0.14859	2.424415	1	1	-1.11505
1.31954	-0.195	1	1	0.490365

第2章 中央競馬の仕組み

中央競馬

1954年に日本中央競馬会(JRA)が発足され、競馬法をもとに中央競馬が開催されるようになった。場所は京都競馬場と東京競馬場の2か所から始まり、今では阪神競馬場、中京競馬場、中山競馬場、新潟競馬場、福島競馬場、札幌競馬場、函館競馬場、小倉競馬場の8か所が加わり、全10会場で順番にレースを開催している。中央競馬ではレースで上位に来る馬を予想して勝ち馬投票券を購入して投票を行う。予想が的中した場合、購入した勝ち馬投票券のオッズに応じて払戻金を受け取ることが出来る。この勝ち馬投票券の売上は次のように構成されている。売り上げの合計が100円であった場合、100円の内75円が購入者に対して払い戻される。残りの25円の内約10円が国庫納付金に割り当てられ、残りの15円が日本中央競馬会の運営費に割り当てられる。仮に運営費を差し引いても利益が残る場合は、利益の半分が国庫納付金に割り当てられる。この時の国庫納付金の合計の4分の1は社会福祉に使用され、残りは畜産振興に対して使用される。

投票方式

競馬には投票方式が複数あり、複雑になっている。その中でも、中央競馬には全部で9つの投票方式がある。単勝、複勝、ワイド、馬単、馬連、枠連、3連単、3連複、win5の9つである。どの投票方式も1票100円であり、それぞれにルールと払戻率が設定されている。まず、単勝はレースで1着になる馬を予想して購入する投票方式である。この投票方式はすべてのレースで発売される。また、設定払戻率は80%となっている。単勝については1頭に対して投票するというシンプルな投票方式であるため、単勝の投票票数に基づいて馬の人気順が設定される。次に、複勝は3着以内、レースによっては2着以内に入る馬を予想する投票方法である。複勝はレースの出走頭数が8頭以上の場合は3着以内、5頭以上7頭以下のレースの場合は2着以内、4頭以下の場合には発売されない。設定払戻率は80%に設定されている。これは単勝に対して的中範囲の広い投票方式であり、1レースに2~3通りの的中があるためそれぞれの勝ち馬投票券の総投票票数に対する割合に応じて払戻は変化する。基本的には単勝よりも少ない払戻になる。次に、ワイドは3着以内に入る馬を2頭選んで2頭とも3着以内である場合に的中となる投票方式である。この投票方式はレースの頭数が4頭以上の場合に発売される。ワイドは1レースで3通りの的中があるので複勝と同様にそれぞれの勝ち馬投票券の総投票票数に対する割合に応じて払戻は変化する。また、設定払戻率は77.5%となっている。次に、馬単は1着の馬と2着の馬を着順を含めて予想する投票方式である。これはレースの出走頭数が3頭以上の場合に発売される。馬単の設定払戻率は75%となっている。次に、馬連は2着以内に入る馬を2頭選び、2頭とも2着以内の場合に的中となる投票方式である。これは馬単と同様にレースの出走頭数が

3頭以上の場合に発売される。また、設定払戻率は77.5%となっている。次に、枠連は馬ではなく、レースのスタート地点にあるゲートの枠番号で2着以内に入る枠番号を2つ選んで投票する方式である。これは1つの枠に対して1~3頭の馬がいるため、馬連に近い投票方式であるが枠連の方がよりの的中しやすくなっている。枠連はレースの出走頭数が9頭以上の場合に発売される。これは枠が1~8枠までであり8頭以下の場合に馬連と全く同じ投票になるためである。また、設定払戻率は77.5%である。次に、3連複は3着以内に入る馬を3頭選んで3頭とも3着以内の場合に的中となる投票方式である。これはレースの出走頭数が4頭以上の場合に発売される。また、三連複の設定払戻率は75%となっている。次に、3連単は1着、2着、3着の順番を含めて予想する投票方式である。これはレースの出走頭数が4頭以上の場合に発売される。また、3連単の設定払戻率は72.5%となっている。最後に、win5は日本中央競馬会が指定する5レースで1着になる馬をそれぞれ予想してすべての1着の馬を的中させる投票方式である。また、win5の設定払戻率は70%となっている。今回の分析では単勝から決まる人気順を利用して分析を行う。その際荒れるレースの指標として1, 2馬の平均人気を用いる。

第3章 分析

今回の分析を行うにあたって日本中央競馬会が提供している JRA-VAN データラボにて公開されている「TARGET frontier JV」から公式のデータを取り込んだ。対象としたデータは東京競馬場で過去 15 年間に行われたレースであるが、障害コースのレースも含まれていた。当然中央競馬のレースの一つではあるが、一般的に行われているのは芝とダートのコースで行われるレースであるので障害コースのレースを除外することを決めた。障害コースのレースを除いた状態でレースは全部で 7477 レースで、これが今回のデータの総数である。また、レース開始前の時点で正式に公開されている情報として「距離」、「馬場」、「出走頭数」、「馬場状況」を取り入れた。このとき、「距離」に関してはメートル単位で表記されることが一般的であるのでそのままメートル単位でデータを取った。「馬場」に関しては前述のとおり障害コースを除いた芝コースとダートコースの 2 つのコースであるのでダミー変数を用いてデータとした。「出走頭数」はレースに登録されていた頭数をデータとした。「馬場状況」に関しては良馬場開催が基本となるため、稍重馬場、重馬場、不良馬場は馬場状況の悪い状態として 1 つにまとめ、良馬場開催かそれ以外の状況の開催かで分けてダミー変数を用いてデータとした。これらを説明変数として R を用いて、重回帰分析を行った。説明変数の中で、「距離」のデータは明らかにほかのデータに対して大きいためダミー変数を除いた説明変数は標準化を行ったうえで重回帰分析をした。

R の分析結果

Call:

```
lm(formula = 平均人気 ~ 距離 + 良 + 芝 + 頭数, data = Tokyo)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.4695	-0.7439	-0.2039	0.5392	5.2188

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.011465	0.023779	0.482	0.62973
距離	-0.003728	0.012316	-0.303	0.76217
良	-0.059147	0.025841	-2.289	0.02211 *
芝	0.064243	0.024428	2.630	0.00856 **
頭数	0.203633	0.011788	17.274	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9795 on 7471 degrees of freedom

Multiple R-squared: 0.04125, Adjusted R-squared: 0.04073

F-statistic: 80.35 on 4 and 7471 DF, p-value: < 2.2e-16

分析の結果から、「距離」は荒れるか否かに影響があるとは言えなかった。一方で、その他の変数である「馬場」、「出走頭数」、「馬場状況」は影響があると判断した。「馬場状況」に関しては良馬場開催であると 1,2 着の平均人気は下がるような結果を示した。「出走頭数」に関しては頭数が多いほど 1,2 着の平均人気は上がるような結果を示した。「馬場」に関しては、芝コースでの開催の場合であると 1,2 着の平均人気は上がるような結果を示した。分析の結果から、「距離」は荒れるか否かに影響があるとは言えなかった。一方で、その他の変数である「馬場」、「出走頭数」、「馬場状況」は影響があると判断した。「馬場状況」に関しては良馬場開催であると 1,2 着の平均人気は下がるような結果を示した。「出走頭数」に関しては頭数が多いほど 1,2 着の平均人気は上がるような結果を示した。「馬場」に関しては、芝コースでの開催の場合であると 1,2 着の平均人気は上がるような結果を示した。

第4章 結論・考察

今回の分析では荒れるレースの要因を 1,2 着馬の平均人気を用いることによって明らかにした。分析の結果から「馬場」はダートコースよりも芝コースで、「出走頭数」はより多い場合で、「馬場状況」は良馬場以外の状況での開催でレースは人気と結果に乖離が生まれるといえる。しかし、今回の重回帰分析では決定係数の値が 0.04125、調整済み決定係数の値が 0.04073 と低いいため回帰モデルとしてはまだ十分に説明できているとはいえない。データ数は十分にあったため、回帰モデルがうまく表せなかった原因として考えられるのは今回使用した変数が少なかったことが考えられる。より精度を上げるには複数の変数を追加する必要があると考えられる。しかし、レース開始前の時点で公開されている情報から荒れるレースに影響がある要素を明らかにしたことは進歩である。

謝辞

本論文を執筆するにあたってご指導いただいた梅津先生に感謝申し上げます。感染症対策のため直接お会いする機会が少なかったものの精神的なサポートを含めて丁寧に対応していただきました。アドバイスや励ましのメッセージを頂いてもまともな返信も出来ず、大変ご迷惑をおかけしました。この未熟な私を見守っていただいたことを改めてお礼申し上げます。

参考文献

- [1] 日本中央競馬会, JRA の概要, 成長の推移
<https://jra.jp/company/about/outline/growth/> (参照 2022/01/10)
- [2] 松田紀之 質的情報の多変量解析 朝倉書店 1988