

中央競馬のオッズに関する統計的分析

2003MM070 室梅秀平

指導教員：木村美善

1 はじめに

「競馬」とは競走馬が決められたコースをいかに早く走破するかを競うスポーツであり、そのレース結果を予想するギャンブルでもある。ギャンブルとしての競馬では、客はレースの結果を予想し、その予想に応じた馬券を買うことでお金を賭ける。見事その予想が当たれば、馬券毎に決められたオッズに応じた払戻金を手にすることが出来る。

「どの馬が勝つか」ということが事前に予想できれば、競馬に負けることは無い。しかし、それは非常に困難である。「どの馬が最も勝つ確率が高いか」ならば予想することが可能かもしれない。しかし、最も勝つ確率が高い馬に賭けることが正解とは限らない。なぜならば、勝つ確率が高い馬ほどオッズが低いことが多いからだ。もし8割の確率で勝つ馬がいたとしても、オッズが1.1倍しかなければ損をする確率のほうが高い。

それならば、「勝つ確率に比べてオッズが高い馬」を予想することは出来ないだろうか。そこで、本研究では主にオッズについて分析を行い、オッズと実際の勝率の関係を調べた。

2 オッズの決まり方

日本の中央競馬のオッズは、パリミュチュエル方式 (Parimutuel betting) と呼ばれる方法で決められている。パリミュチュエル方式とは、賭け金の合計からまず胴元の取り分を引き、その残りを配当金として当選者が分け合う方式である。

全体の賭け金に対し、その馬に対して賭けられている金額が多いほどオッズは小さくなり、逆にその馬に対して賭けられている金額が少ないほどオッズは大きくなる。この「全体の賭け金に対して、その馬に賭けられた金額の割り合い」をその馬の「支持率」と呼ぶ。各馬のオッズはその馬の支持率によって決定されるため、支持率が分かればオッズは算出可能である。また、逆にオッズから支持率を算出することも可能である。具体的な計算式は次のようになる。(JRA ホームページ [1] より)

その馬の支持率を a 、オッズを b として

$$\frac{0.788}{a} + 0.1 = b$$

オッズの算出には支持率が使用されるが、最終的な支持率は馬券の購入が締め切られるまで確定しない。しかし、締め切り前の時間帯であっても、その時の支持率から暫定オッズが算出され、公表されている。馬券購入者達は、その暫定オッズを参考にしながら購入馬券を検討することができるようになっている。

たいていの馬券購入者は、馬券を買うときにオッズを参考にしている。もし、彼らが妥当と考えるオッズよりも暫定オッズが高い場合、その馬券は売れ、暫定オッズ

は下がっていくだろう。逆に、彼らが妥当と考えるオッズよりも暫定オッズが低い場合、その馬券は売れず、暫定オッズは上っていくだろう。そのように考えると、「支持率」というものは「馬券購入者達が考えている勝率」に近づいていくと考えられる。

3 分析方法

前述の通り、「支持率」が「馬券購入者達が考えている勝率」を表すと考えれば、「支持率」と「実際の勝率」の関係を分析すれば「馬券購入者達の考えている勝率」と「実際の勝率」との間のギャップを調べることが出来る。しかし、各馬の「支持率」はそれぞれ分かっているのに対し、「実際の勝率」は分からない。勝利したか敗北したかという結果のみが分かっている。

そこで、分析方法としてロジスティック回帰分析を選択した。

4 ロジスティック回帰分析

説明変数群 (x_1, x_2, \dots, x_n) を用いて、確率のような範囲 $(0,1)$ の変数を説明しようとする場合、通常重回帰分析では説明変数群の合成変数の取りうる値の範囲が $(-\infty, \infty)$ となってしまうため、適当ではない。

そこで、下式のように説明変数群の合成変数をロジスティック関数にすることで、その値域を範囲 $(0,1)$ に収めることができる。(参考文献 [2] を参照のこと)

$$y = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}$$

従属変数には勝敗を、説明変数には支持率を使用し、支持率によって勝敗を説明する形を基本とする。そして、説明変数に適当な変数を追加することで、その変数が支持率と勝敗の関係をどのように変化させるかを調べようと考えた。

5 使用したデータ

競馬予想に詳細な情報は必須である。JRA(日本中央競馬会)も色々と詳細なデータを提供しているし、インターネット上で無料で手に入る分だけでも豊富なデータをそろえることが可能である。

R や Excel への転記の容易さなどから、インターネット上の競馬情報サイト「netkeiba.com[3]」を選び、データは主にそのサイトと JRA の公式ホームページ [1] から収集した。

データは、2008年に中央競馬において行われた全3452レースとその出走馬延べ50215頭用意した。

6 分析例:1番人気について

例を兼ねて、「1番人気」という要素について分析を行ってみたい。まず、単純に「1番人気」という要素のみを説

明変数としてロジスティック回帰分析を行う。なお、「1番人気」は1番人気を1、それ以外を0とするダミー変数である。

表 1: 1番人気についての分析結果 (1)

	回帰係数	標準偏差	t 値	p 値
定数項	-2.95582	0.02133	-138.60	<2e-16 ***
1番人気	2.25135	0.04199	53.61	<2e-16 ***

分析結果を見ると、「1番人気」は正の方向に強く効いている。「1番人気」の馬はそうで無い馬に比べて勝率が高いという分析結果であるが、馬券購入者達が、最も勝率が高いと考えている馬と考えているのが1番人気になるため、実際の勝率も当然高い。本研究にて興味があるのは「その要素が勝率に影響を及ぼすか」ではなく、「その要素が支持率と勝率の関係にどのような影響を及ぼすか」である。そこで、説明変数に「支持率」を取り入れる。

ロジスティック回帰分析では説明変数群をロジスティック変換して値域(0,1)の確率変数の形とする。そこで、既に確率変数の形である「支持率」は、ロジスティック変換した後の形が確率変数の形を取るようにロジスティック関数の逆関数(下式)で変換しておく。

支持率を a 、変換後の支持率を A として

$$A = \log \frac{a}{1-a}$$

このように変換した「支持率」と、先程使用した「1番人気」を説明変数として、改めて分析を行ってみた。

表 2: 1番人気についての分析結果 (2)

	回帰係数	標準偏差	t 値	p 値
定数項	0.02410	0.05635	0.428	0.6688
支持率	1.00721	0.02155	46.742	<2e-16 ***
1番人気	-0.09771	0.05837	-1.674	0.0941

先程と比較すると、「1番人気」の回帰係数が正負逆転している。あまり強くは無いが負の方向に効いているということは、1番人気の馬は支持率に比べて勝率が低い、または、実際の勝率に比べて支持率が高いという分析結果となる。この場合、1番人気の馬は人気が過大になりやすく、オッズの割には勝率が低くなり易いと言える。

7 騎手に関する分析

本研究にて行ったいくつかの分析の内、最も影響が大きかった「騎手」についての分析結果の一部をここに記載する。

騎手の分析では、対象の騎手が乗った馬を1、それ以外の馬を0とするダミー変数を作成し、それを説明変数に加えて分析を行う。中央競馬において、2008年度のレースに騎乗した騎手は247人に上る。本研究では、この内20人に絞って分析を行ったが、ここでは2008年の最多勝利騎手である「武豊」に絞る。なお、各分析結果において「定数項」と「支持率」の部分には大きな違いが無い。そのため、紙面の節約のために以下では「定数項」と「支持率」は割愛させて頂いた。

まず、全レースを対象に分析を行う。

表 3: 武豊についての分析結果

	回帰係数	標準偏差	t 値	p 値
武豊	-0.09397	0.10519	-0.893	0.372

騎手「武豊」という要素はあまり効いていない。次に、全レースをクラスごと(新馬戦、未勝利戦、条件戦、オープン戦)に分けて、それぞれに同様の分析を行うと、それぞれ以下ようになる。

表 4: 武豊についての分析結果 (クラス別)

	回帰係数	標準偏差	t 値	p 値
武豊(新馬)	-0.26162	0.34686	-0.754	0.451
武豊(未勝利)	0.32204	0.17977	1.791	0.0732
武豊(条件)	-0.20740	0.15666	-1.324	0.186
武豊(オープン)	-0.62389	0.35039	-1.781	0.0750

未勝利戦においては正の方向に効いているにも拘らず、オープン戦では負の方向に効いている。武豊騎手はオープン戦では実際の勝率より過剰な支持率となり易く、逆に条件戦では支持率以上に勝ち易いと言える。

次に、レースを距離毎に分けて同様の分析を行う。

表 5: 武豊についての分析結果 (距離別)

	回帰係数	標準偏差	t 値	p 値
武豊(短距離)	-0.10645	0.23140	-0.460	0.645
武豊(マイル)	-0.02889	0.13192	-0.219	0.827
武豊(中距離)	-0.91606	0.39017	-2.348	0.0189 *
武豊(長距離)	0.27639	0.38730	0.714	0.475

中距離(1900m~2100m)において、支持率が過剰になり易いという結果となった。この距離帯のレースには大レースが多いことや、オープン戦についても同様に負の方向に効いていたことから、武豊騎手は特に大きいレースにおいて過剰な支持率を得やすいと考えられる。

8 おわりに

実際の勝率に比べてオッズが高くなる状況があるとして、そのようなケースに該当する馬に賭け続ければ、仮に的中率が低くてもトータルでプラスになるのではないだろうか。

本研究はそのような考えから思い至ったものだが、実際には、オッズが勝率より少し高いくらいではプラスにならなかった。競馬には約25%の控除(胴元の取り分)があるからである。

また、本研究において分析に用いた各要素を比べると、騎手に関する要素の影響が大きい傾向にあった。騎手の能力という要素について、馬券購入者達のイメージと実際の騎手の実力に違いがあることが多いように感じられた。

参考文献

- [1] JRA(日本中央競馬会) 公式ホームページ <http://www.jra.go.jp/index.html>
- [2] 中村永友著、金明哲編、「Rで学ぶデータサイエンス 2 多次元データ解析法」、共立出版、2009
- [3] 競馬予想サイト [netkeiba.com](http://www.netkeiba.com) <http://www.netkeiba.com/>