

競走馬の斤量差に関する統計的分析

2010SE213 鈴木駿太

指導教員：松田真一

1 はじめに

近年の日本競馬では牝馬（メス）が牡馬（オス）と同じレースを走り、好成績を残すことが多くなった。しかし牝馬限定戦の存在が示すように、以前は牝馬は牡馬相手ではほとんど歯が立たなかった。元々牝馬には牡馬と一緒に走るときには能力差があると考えられ、斤量面でアドバンテージが与えられている。しかし近年牝馬の活躍が目立つことで斤量のアドバンテージは本当に適切なのかということが気になり、そういった趣旨から斤量差に関する分析をしたいと思います。

2 斤量の決定

現在日本の競馬のレースでは主に牝牝の斤量差を2キロとしている。この差は競走馬のレースのデビュー時期である馬齢が2歳だと1キロ差だったり、斤量差がなかったりする。出走馬すべてに勝つチャンスを与えようという趣旨のハンデ戦のように人間の主観によって斤量が決定される場合でも、この牝牝の2キロ差を考慮したうえで決めている。今回の研究の主なポイントは牝牝の斤量差が2キロであるのが適切かどうかを分析することである。

3 使用するデータ

JRA 公式サイト [1] や各競馬情報サイトにも豊富なデータがあるが、Excel への転記の容易さから Yahoo! Japan の競馬カテゴリー [3] にあるデータを用いる。競馬の世界では拮抗したレースが展開されるように競走馬のレベルに合わせたクラス分けがなされている。本研究では低レベルのレースでは分析精度が悪くなると判断し、最上級のオープンクラスを主な分析対象とした。

4 分析方法

今回は競馬場や距離などレースのそれぞれの条件から出される各競走馬のタイムがどの因子によるものが強いかを分析したいので重回帰分析を用いる。

5 分析手順

目的変数は与えられた斤量などの影響からそれぞれの競走馬の走破能力を予測したいので「タイム」とする。説明変数として用意するのは、今回重視したいポイントである「性別」や「斤量」は勿論のこと「枠番」「馬齢」「騎手」「馬体重」「季節」「馬場状態」の計8項目である。これを競馬場やコース（芝・ダート）、設定距離ごとに分析をかけ牡馬と牝馬の走破能力（タイム）に差があるかどうか、それは性別や斤量が影響しているのかどうか、他の因子が影響しているのかを発見したい。

この説明変数の中で、「性別」「騎手」「季節」「馬場状態」

はダミー変数として扱う。「性別」の項目は競走馬の性別として去勢されたせん馬もあるが、斤量設定は牡馬と同じであり牡馬や牝馬と比べ圧倒的に数が少ないので本研究では牡馬としてカウントする。よって性別は牡馬か牝馬の2択となるため「性別牡」という説明変数を用意する。また「騎手」の項目は各競走馬に騎乗しうる騎手が100人以上と多数おり1人1人因子を用意しても良い分析は行えないと判断し、netkeiba[2]のデータを使用し各騎手の連対率を基に4段階で騎手ランクを作成した。分析の結果、性別や斤量の因子の影響が強いならば性差による斤量の補正が効いていないということができる。分析精度向上のためデータは3年分（2011年～2013年）の当該レース条件の全レースのデータで分析する。また近年の牝馬の活躍を示したいので10年前（2001年～2003年）の同レース条件での分析比較も行う。

6 分析結果

6.1 近年と10年前の分析結果

今回は紙面の都合上1つのレース条件のみ取り上げて分析する。オープンクラスで近年最も多く施行されているレース条件が京都芝1200mである。斤量設定に関してはハンデ戦もしくは別定戦のデータである。この条件の分析対象は18レース、分析データ数280である。また季節は春、秋、冬で実施されたデータがある。ゆえに説明変数として「季節春」「季節秋」を用意しダミー変数として扱った。馬場状態は良、やや重、重、不良のデータがある。ゆえに説明変数として「馬場状態やや重」「馬場状態重」「馬場状態不良」を用意しダミー変数として扱った。

Rのstepコマンドで変数選択を行ったところ表1のような結果になった。

表1 変数選択後の分析結果（京都芝1200m）

	回帰係数	標準誤差	t 値	P 値
(定数項)	65.8077	1.7258	38.132	0.0000
枠番	0.0256	0.0165	1.551	0.1220
性別牡	-0.2260	0.1042	-2.170	0.0309
馬齢	0.0789	0.0278	2.838	0.0049
騎ランク d	0.3152	0.1083	2.910	0.0039
斤量	0.0513	0.0327	1.572	0.1171
季節春	-1.0751	0.1432	-7.507	0.0000
季節秋	-0.7274	0.0951	-7.646	0.0000
馬場やや重	0.5807	0.1740	3.338	0.0010
馬場重	1.8434	0.1731	10.652	0.0000
馬場不良	2.1441	0.1913	11.207	0.0000

決定係数:0.6203 自由度調整済決定係数:0.6062

この条件では枠番, 性別, 馬齢, 騎手, 斤量, 季節, 馬場状態の7つの因子がタイムを予測する上で重要な因子といえ, 牝馬での斤量の能力差の補正は効いているとはいえないことを示している. 今回の研究で注目したいのは牝馬と牡馬の間に設けられている斤量差は妥当かどうかということなので着目すべきは「性別牡」と「斤量」である. 「性別牡」の回帰係数が負であるので牝馬のほうが速く走れる, すなわち牝馬による能力差はあるといえる. 「斤量」の回帰係数は正であるので斤量が重いほど走破能力が落ちるといえる. 具体的な数値を見ると性別に関しては牝馬の方が0.20秒速く走れ, 斤量に関しては2キロ減ると $2 \times 0.05 = 0.10$ 秒速く走れるといえる. よって2キロの斤量差を設けても牝馬は牡馬に対して約0.10秒走破能力が劣るといえる. ちなみに0.10秒は着差としておよそ半馬身である.

次に残差分析を行うために4つの図を出力する. 図1の左上の残差をみる図と左下の標準化残差をみる図を見ると, プロットが散らばっており一定の傾向がないことから分散の均一性を確認できる. 右上の正規性をみる図からプロットが直線上にあるので誤差の正規性を確認できる. 右下の誤差の外れ値をみる図からは外れ値の存在を確認できない. よって図1より重回帰分析が使えるモデルであると判断できる.

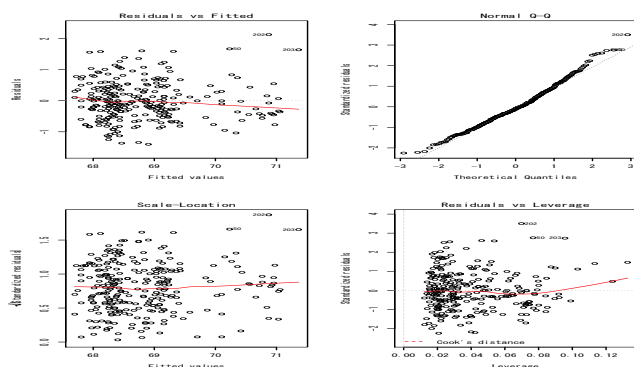


図1 回帰診断図(京都芝1200m)

VIF(分散拡大要因)による多重共線性の有無を確認すると変数選択後に残る変数のうちVIF値が最高値となる変数の「斤量」でもVIF値が2.0815となり, 変数選択後に残った変数でVIF値 > 10 となる変数は存在しないので多重共線性の疑いは低いといえる.

10年前の同レース条件では分析対象9レース, 分析データ数123であり同じように分析した. そうしたら10年前は変数選択で「性別牡」も「斤量」も残らなかった. ゆえに斤量差によって能力差は補正できていたと考えられる.

6.2 適切な斤量差の考察

このレース条件が近年中央競馬で行われたオープンクラスで一番多いデータ数であり, 当然この条件で何頭かの馬

が何度も走っている. そこでこの条件で多く走っていて実力が拮抗していると思われる特定の4頭の馬を取り上げ, 斤量について詳細に考察していく. この4頭すべてもしくは一部の馬を使って5パターンで重回帰分析を行ったところ全てで変数選択において「斤量」の変数が残った. その中で斤量の回帰係数の値やP値を比較し, 一番安定していると判断した特定の3頭を利用して分析した結果を利用する.

「斤量」の変数のP値や決定係数の値から近年全体で分析したものよりこちらの「斤量」の回帰係数の値を支持したいと考える. 近年全体の「性別牡」の回帰係数 -0.22603 と特定の3頭から導かれた「斤量」の回帰係数 0.4674 から $-0.22603/0.4674 = -0.48359$ であるので, およそ5キロ斤量差を大きくすると, 性差の補正はできることになる. よってこの条件では牝馬には牡馬より2.5キロ軽い斤量を与えると牡馬と対等に戦えるといえる.

7 まとめ

本研究では斤量差について考察し, 牝馬が斤量のアドバンテージをもらい過ぎる条件はあるのかという趣旨から始めたものだが, 2歳戦も含めデータサンプルの多い計16のレース条件で分析を行ったが一つも牝馬が有利となっている条件はなかった. 近年の方が10年前より牡馬混合G1を勝った牝馬が多いので近年牝馬が活躍していることは否定することはないが, G1の中でも賞金の高いレースで活躍する牝馬の影響で牝馬全体が活躍する印象だった. しかし今回の分析結果を基にするとそれは特定の馬の大舞台での活躍が目立つからであるという結論を得た. 性別の変数が変数選択で消える条件の方が多く, 現行の斤量設定は正しいといえそうである. また性別の変数が残る条件でも全て牝馬有利の回帰係数の符号となった. このことから全体的に斤量差を現行の2キロから縮める必要は無く, 牝馬有利な条件で斤量差を大きくしようにも京都芝1200mのように最大でも2.5キロ差までいいという結論が出た.

8 おわりに

牝馬が牡馬を蹴散らして活躍するシーンが近年目立つ中, 牝馬に有利な斤量設定のレース条件が無かったことには少し驚いた. ただ本研究によって牝馬の全体数が微増していることと特定の馬の活躍でそう見えていたという発見ができてよかったと思う. 本研究を終えて, 私は競馬に関心が強いので競馬に関する統計的分析に今後またチャレンジしてみたい気になった.

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/>
- [2] netkeiba.com : <http://www.netkeiba.com/>
- [3] Yahoo!JAPAN スポーツ 競馬 :
<http://keiba.yahoo.co.jp/>