卒業論文

主成分分析と因子分析による競馬の勝因の研究

大阪工業大学
情報科学部　情報科学科
学生番号　A04-133
辺見　広大

平成21年2月9日

目 次

1　序論

1.1　背景

競馬とは、騎手と馬が一定の区間をどれだけ速く駆け抜けられるかを競うスポーツである。正式のルールに基づき、専用の競技用施設において行われる競馬は、１６世紀のイギリスから始まった。日本ではＪＲＡ(日本中央競馬会) が主催・施行を行なっている。ＪＲＡが主催で行なわれる競馬で走る馬は、品種改良された軽種馬であるサラブレッドが用いられる。

全国各地にＪＲＡが主催で行なうコース形態の異なった競馬場が１０箇所あり、競馬場によって右回りの、左回りのが決まっている。走るコースは、芝コース、ダート（砂）コース、障害のコースがある。日本では主に芝のコースが主流である。馬の強さを分けるために獲得賞金に応じて決められる。その中でも格の高いレースを重賞競争（グレードレース）と呼び格の高い順番にＧⅠ、ＧⅡ、ＧⅢとなる。基本的には馬と騎手がメインのスポーツであるが、他に係わる人として、馬を生産する生産者、馬を調教する調教師と調教助手、馬の世話をする厩務員、馬の持ち主である馬主がいる。

競馬はスポーツの他にギャンブルとしての意味合いも強い。行なわれるレースの１着、２着、３着を勝ち馬投票券（以後馬券と書く）を買って当てるというものである。馬券には種類があり、１着を当てる単勝、１～３着の内どれか１頭を当てる複勝、１着、２着の枠を当てる枠連、１着、２着を当てる馬連、１～３着の内どれか２頭を当てるワイド、１～３着を当てる３連複、１着、２着を順番通り当てる馬単、１～３着を順番通り当てる３連単がある。１～３着に入ることが馬券に絡むということになる。その日に行なわれるレースと、レースに走る馬の情報がスポーツ新聞、競馬新聞に載っている (以後総称して競馬新聞と書く)。競馬新聞に載っている馬の成績を見て馬券を買う時の参考にする。

競馬新聞には馬の右回り成績、左回り成績、競馬場の成績、距離成績、馬の近走の成績等多くのデータが載っている。馬券を買う人はそのうち何かのデータを元にするが、本研究では統計的に多く使われるデータを用いて進めた。出走している馬がレースに勝利する要因として、勝ち馬にどの成績が影響して勝っているのかを調べた。本研究の最終目標は、実際のレースに当てはめて、研究の結果の勝ち馬を出し、実際のレースでその馬が勝つのかを調べた。

1.2　本研究の内容

主成分分析と因子分析を用いて、中山競馬場の２５００ｍで行なわれる有馬記念というＧＩレースを予想した。３要素を読み込み、主成分分析と因子分析を計算するプログラムをＣ言語で作成した。

論文中で詳しく書くが、簡単に主成分分析と因子分析の説明をする。

主成分分析とは、多数の要素についての統計データが与えられたとき、それらの１次結合で表現される新たな要素を構成し、最終的には、観測結果をもとの変量の個数よりも少ない個数の変量の動きにまとめあげようというものである。

因子分析とは、実際のわかっている要素を、$y_{1i} = b_{11}f_{1i} + b_{12}f_{2i} + e_{1i}$ であると仮定する。i は対象のデータの識別番号を表す。b_{11}、b_{12} は要素によって変わるが、同じ要素の別の対象のデータでは変わらない。f_{1i}、f_{2i} 要素では変わらないが、対象のデータによって変わる数である。この b や f や e の値を統計的に推定する方法である。

研究方法としては、過去４年分の中山競馬場２５００mの勝ち馬４２頭の右回り成績、左回り成績、競馬場の成績、距離成績と、過去４年分の有馬記念の出馬６１頭の右回り成績、左回り成績、競馬場の成績、距離成績、GＩの成績、重賞競争の成績から３変量を選んで主成分分析と因子分析を行なう。成績の選び方としては、馬に対しての成績を使用したため、馬の成績のみで分析する。それぞれ理由としては、右回り成績、左回り成績は人間にもあるように、馬にも利き手があるため、それを調べるために使用し、競馬場の成績は競馬場によってコース形態が違うため、それを調べるために使用し、距離成績はその距離をどれだけ走ったことがあり、なおかつ得意かを調べるために使用し、GＩの成績、重賞競争はその馬の強さを決めるために用いた。

過去４年分の中山競馬場２５００mの勝ち馬のデータを元に、利き手やコースや距離がどの用に勝ち馬に影響しているのか調べるために用いた。過去４年分の有馬記念の出馬のデータは利き手、コース、距離、馬の格がどのように勝ち馬に影響しているかを調べる。

データは１着、２着、３着、着外の成績を用いて主成分分析と因子分析を行う。なお変量は標準化した値を用いる。３変数は上記のデータを組み合わせて議論する。

基本は馬券圏内である１着、２着、３着に重みをつけて変数を決める。上記のデータが種のデータの相関を調べ、勝ち馬との対応を分析する。

1.3 本研究論文の構成

本論の構成は以下の通りである。まず第２章で、主成分分析と変量の標準化について説明する。第３章で因子分析の説明をする。そして第４章で今回行なった分析の方法を詳しく説明し、それ以降の章では分析の、結果と考察を主成分分析と因子分析の比較を交えながら述べる。

2　主成分分析

この章では主成分分析について説明する。尚、この章では、文献 [1] を参考にした。なお変量とは対象となる馬の要素のことであり、個々に異なった値をとる。

2.1　合成変量の構成

主成分分析とは、多数の変量についての統計データが与えられたとき、それらの１次結合で表現される新たな変量を構成し、最終的には、もとの変量の数よりも、少ない変量の動きにまとめあげようというものである。

ある程度同質性があると思われる二つの変量を両軸にとって、対象の集団に属する個々の対象の要素がどのように分布しているかを散布図で描けば、必ずしも 45 °の方向とは限らないが、主要な傾向として、左下から右上へ向かう線に沿った並び方が見だされる。

その散布図を仮に楕円と考えた場合、その長径に平行な方向と直角な方向に新たな座標軸をとって、個々の点の座標をこの座標系で読むことにする。主成分分析で用いる合成変量とは長径方向の座標や短径方向の座標に相当する数量として構成される。構成の方法は、２次元での場合、元の変量が x_1、x_2 であるとき、

$$z = a_1x_1 + a_2x_2 \tag{2.1}$$

によって合成変量を定義し、係数について

$$a_1{}^2 + a_2{}^2 = 1 \tag{2.2}$$

という制約を課す。そうすると、この z は、x_1 と x_2 を成分とするベクトル $\boldsymbol{x}$ の、ある直線の方向への正射影の長さであるといえ、(2.2) 式が満たされるとき、a_1 と a_2 は何らかの角度の余弦と正弦になる。つまり、

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{2.3}$$

である。したがって、ベクトル $\boldsymbol{a}$ は原点を通って横軸となす角 θ であるような直線 l の方向ベクトルである単位ベクトルとなる。

$$z = a_1x_1 + a_2x_2 = \langle \boldsymbol{a}, \boldsymbol{x} \rangle \tag{2.4}$$

は、ベクトル $\boldsymbol{x}$ を直線 l の方向に射影した正射影の長さになる。

2.2　合成変量の分散の最大化

以下では対象のデータが n 個である。楕円の長径方向にあたるような z_1 軸、同じことだがベクトル $\boldsymbol{a}$ を、決めるために、元のデータを $x_{1i}, x_{2i}(i = 1, 2, \cdots\cdots, n)$ として、それらを (2.1) 式にあてはめたとき得られる値を $z_i(i = 1, 2, \cdots\cdots, n)$ とする。z_i 全体について分散を求め、それが (2.2) 式のもとで最大になるように a_1、a_2 を決めると、長径の方向が求まる。

合成変量 z の分散は

$$
\begin{aligned}
Var(z) &= \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 \\
&= {a_1}^2 v_{11} + 2a_1 a_2 v_{12} + {a_2}^2 v_{22}
\end{aligned} \tag{2.5}
$$

となる。(2.5) 式の $\bar{z}$ は合成変量 z の平均値、v_{11}、v_{22} は変量 x_1、x_2 の分散であり、v_{12} は変量 x_1 、x_2 の共分散である。(2.5) 式ラグランジュ乗数法を用いてラグランジュ関数を構成すると、

$$
L(a_1, a_2, \lambda) = {a_1}^2 v_{11} + 2a_1 a_2 v_{12} + {a_2}^2 v_{22} + \lambda(1 - {a_1}^2 - {a_2}^2) \tag{2.6}
$$

となる。これを 3 つの変数について偏微分して整理すると

$$
v_{11}a_1 + v_{12}a_2 = \lambda a_1 \tag{2.7}
$$

$$
v_{21}a_2 + v_{22}a_2 = \lambda a_2 \tag{2.8}
$$

$$
{a_1}^2 + {a_2}^2 = 1 \tag{2.9}
$$

ここで (2.7) 式、(2.8) 式は、

$$
V = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}, \quad \boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}
$$

でおくと、

$$
V\boldsymbol{a} = \lambda\boldsymbol{a} \tag{2.10}
$$

と簡単に書き直せる。行列 V は変量 x_1 と x_2 の分散共分散行列とよばれる。$v_{12} = v_{21}$ であり

$$
Var(z) = \lambda \tag{2.11}
$$

である。(2.10) 式は行列 V の固有値 λ と固有ベクトル $\boldsymbol{a}$ を (2.2) 式の制約のもと求める問題に相当する。先に述べたとおり、分散が最大になるような値を求めればいいので、(2.11) 式より、λ の値は 2 つ出てくるので、大きいほうの値をとり、λ_1 とし、それに相当する固有ベクトル $\boldsymbol{a_1}$ を求める。

このベクトル $\boldsymbol{a_1}$ によって導きだされる合成変量

$$
z_1 = a_{11}x_1 + a_{21}x_2 \tag{2.12}
$$

を第 1 主成分とよぶ。個々の対象のデータをこの式にあてはめたときに得られる z_1 の値

$$
z_{1i} = a_{11}x_{1i} + a_{21}x_{2i} \tag{2.13}
$$

のことを、対象のデータに関する主成分得点という。

変量が n でも成り立つため、変量が 3 個の場合は、行列 V が 3×3 の分散共分散行列になるので、固有値と固有ベクトルが 3 個出る。

2.3　寄与率

§2.2 で述べたとおり変量が n 個の場合、固有値 λ と固有ベクトル $\boldsymbol{a}$ は n 個出てくる。ここで

$$\mu_i = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \cdots + \lambda_n} \tag{2.14}$$

という指標を定義し、第 i 主成分の寄与率とよぶ。μ_i はもとの変量の全変動の何パーセントが第 i 主成分の変動で説明できるか示す指標である。μ_i の値が大きいほどもとの変量の多くを束ねているといえる。さらに、

$$\nu_i = \sum_{k=1}^{i} \mu_k \tag{2.15}$$

という指標を考え、第 i 番目までの主成分の累積寄与率とよぶ。これは、第 1 番目から第 i 番目までの主成分によって全体の変動の何パーセントか説明できる。

あらかじめ約 80 %と決めておき、累積寄与率がその数値を超えたら主成分の採用を打ち切るという形で用いる。

2.4　標準化

2.4.1　変量の標準化

変量の標準化とは、もとの変量の平均値から偏差をとり、それを標準偏差で割ることをいう。もとの変量が x_1、x_2 であるなら、それを標準化した新変量 y_1、y_2 は、

$$y_1 = \frac{x_{1i} - \bar{x_1}}{\sqrt{Var(x_1)}} \tag{2.16}$$

$$y_2 = \frac{x_{2i} - \bar{x_2}}{\sqrt{Var(x_2)}} \tag{2.17}$$

によって定義される。

2.4.2　新変量の分散共分散行列

これらの変量の平均である $\bar{y_1}$、$\bar{y_2}$ はともに 0 である。y_1 の分散は、

$$\begin{aligned} Var(y_1) &= \frac{1}{n}\sum_{i=1}^{n}(y_{1i} - \bar{y}_1)^2 = \frac{1}{n}\sum_{i=1}^{n}{y_{1i}}^2 \\ &= \frac{1}{n}\frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}{Var(x_1)} = \frac{Var(x_1)}{Var(x_1)} = 1 \end{aligned} \tag{2.18}$$

である。y_2 についても同様なので、新変量の分散はともに 1 である。y_1、y_2 の間の共分散は、

$$
\begin{aligned}
Cov(y_1, y_2) &= \frac{1}{n}(y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2) = \frac{1}{n} y_{1i} y_{2i} \\
&= \frac{1}{n} \frac{\displaystyle\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{Var(x_1)}\sqrt{Var(x_2)}} \\
&= \frac{Cov(x_1, x_2)}{\sqrt{Var(x_1)}\sqrt{Var(x_2)}} \equiv r(x_1, x_2) \qquad (2.19)
\end{aligned}
$$

つまり、新変量のあいだの共分散は、もとの変量のあいだの相関係数である。新変量のあいだの相関係数もこれと同じである。結論として、行列 V は

$$
V = \begin{bmatrix} r(x_1, x_1) & r(x_1, x_2) \\ r(x_2, x_1) & r(x_2, x_2) \end{bmatrix} = \begin{bmatrix} r(y_1, y_1) & r(y_1, y_2) \\ r(y_2, y_1) & r(y_2, y_2) \end{bmatrix} \qquad (2.20)
$$

となり、もとの変量のあいだの相関行列であり、同時に新変量のあいだの相関行列になる。これは、相関係数は単位のとり方に依存しないという性質をもつためである。そこで、これを簡単に

$$
V = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \qquad (2.21)
$$

と書く。主成分分析の変量を標準化し、分散共分散行列を相関行列に置き換えて議論する。標準化することにより、もともと分散の値と平均の値がの違うものを、分散の値を 1 平均の値を 0 にして、単位や分布に関係なく議論を進めることができるようになる。

主成分分析の計算方法は、標準化した変量を用い、相関行列を元に、固有値を (2.11) 式で求め、固有ベクトルを計算し、(2.13) 式で主成分得点を計算した。

3 因子分析

この章では因子分析について説明する。変量は標準化をしているため、分散共分散行列は相関行列を用いるので、V は相関行列を表す。3 変量、2 因子を用いるため、論文中も 3 変量、2 因子の場合を書く。尚、この章では、文献 [1] を参考にした。

3.1 因子分析の構成

標準化した変量を y_1、y_2、y_3 とする。対象となるデータを n 個とし、個々についての変量を y_{1i}、y_{2i}、$y_{3i}(i = 1, 2, 3, \cdots, n)$ とする。この３つの変量に影響を与える陰の要因は基本的には 2 種類でつくされていると仮定するなら、y_{1i}、y_{2i}、y_{3i} の具体的な変量の値を説明する仮説の式として、

$$\left.\begin{aligned} y_{1i} &= b_{11}f_{1i} + b_{12}f_{2i} + e_{1i} \\ y_{2i} &= b_{21}f_{1i} + b_{22}f_{2i} + e_{2i} \\ y_{3i} &= b_{31}f_{1i} + b_{32}f_{2i} + e_{3i} \end{aligned}\right\} (i = 1, 2, \cdots\cdots, n) \tag{3.1}$$

と表すことができる。f_{1i}、f_{2i} は、陰の要因として想定される変量 f_1、f_2 の、元の変量 y_1, y_2, y_3 についての具体的な値を示しており、因子得点という。これは対象となるデータによって異なった値をとるため f_{1i}、f_{2i} となる。この陰の変量 f_1、f_2 は共通因子とよぶ。y_1、y_2、y_3 が平均 0 を中心に分布する変量となっているため、右辺の変量は、平均 0 である。さらに f_1、f_2 については、標準偏差が１であると想定する。この陰の変量 f_1、f_2 は共通因子と呼ばれる。b_{11}、b_{21}、b_{31} 、b_{12}、b_{22}、b_{32} の係数は因子負荷量とよばれ、変量によらない定数と想定される。例えば b_{32} は、3 個目の変量に第２因子がどれだけ反映するかの影響度を表す。なお、共通因子の因子得点は対象のデータにより異なる値をとるが、変量によっては変わらないことに注意が必要である。

最後に、以上では説明しきれない部分を表すものとして、独自因子とよばれる変量 e_1、e_2、e_3 と置く。元の変量 y_1, y_2, y_3 については、e_{1i}、e_{2i}、e_{3i} だが、変量ごとに異なる値をとる。e はともに平均 0、標準偏差は異なった値をとると想定し、d_1、d_2、d_3 で表す。

e_{1i}、e_{2i}、e_{3i} はお互いに無相関 (共分散が 0)、それらと f_1、f_2 とのあいだも無相関と仮定する。このときに得られる解を直交解といい、この場合の解として求められた因子を直交因子とよぶ。

3.2 相関行列の分解

１次結合によって表された変量の分散は以下のことが成り立つ。

$$Var(ax + by) = a^2Var(x) + 2abCov(x, y) + b^2Var(y) \tag{3.2}$$

ここで x と y が無相関のとき

$$Var(ax + by) = a^2Var(x) + b^2Var(y) \tag{3.3}$$

となる。

２種の１次結合の間の共分散は次のようになる。

$$\begin{aligned}
&Cov(ax+by, cx+dy) \\
&\qquad = \frac{1}{n}\sum_{i=1}^{n}\{a(x_i-\bar{x})+b(y_i-\bar{y})\}\{c(x_i-\bar{x})+d(y_i-\bar{y})\} \\
&\qquad = acVar(x)+(ad+bc)Cov(x,y)+bdVar(y) \qquad (3.4)
\end{aligned}$$

ここで x と y が無相関のとき

$$Cov(ax+by, cx+dy) = acVar(x)+bdVar(y) \qquad (3.5)$$

となる。

ここで因子分析が直交解の場合、変量 y_1, y_2, y_3 の分散と共分散について、以下のような計算が成り立つ。

$$\begin{aligned}
Var(y_j) &= Var(b_{j1}f_1+b_{j2}f_2+e_j) \\
&= {b_{j1}}^2Var(f_1)+{b_{j2}}^2(f_2)+Var(e_j) \\
&= {b_{j1}}^2+{b_{j2}}^2+{d_j}^2 \quad (j=1,2,3) \qquad (3.6) \\
Cov(y_j,y_k) &= Cov(b_{j1}f_1+b_{j2}f_2+e_j, b_{k1}f_1+b_{k2}f_2+e_k) \\
&= b_{j1}b_{k1}Var(f_1)+b_{j2}b_{k2}Var(f_2) \\
&= b_{j1}b_{k1}+b_{j2}b_{k2} \quad (j=1,2,3 \ \ k=1,2,3 \ \ j\neq k) \qquad (3.7)
\end{aligned}$$

なお j、k は変量の個数である。そこで、y_1, y_2, y_3 の相関行列は、

$$V=\begin{bmatrix} {b_{11}}^2+{b_{12}}^2+{d_1}^2 & b_{11}b_{21}+b_{12}b_{22} & b_{11}b_{31}+b_{12}b_{32} \\ b_{21}b_{11}+b_{22}b_{12} & {b_{21}}^2+{b_{22}}^2+{d_2}^2 & b_{21}b_{31}+b_{22}b_{32} \\ b_{31}b_{11}+b_{32}b_{12} & b_{31}b_{21}+b_{32}b_{22} & {b_{31}}^2+{b_{32}}^2+{d_3}^2 \end{bmatrix} \qquad (3.8)$$

の形で表せる。(3.1) 式より、縦のひとつひとつの列をベクトルとみて

$$\boldsymbol{b}_1=\begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}, \boldsymbol{b}_2=\begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} \qquad (3.9)$$

として、また、${d_1}^2$、${d_2}^2$、${d_3}^2$ を対角成分とする対角行列を

$$D=\begin{bmatrix} {d_1}^2 & 0 & 0 \\ 0 & {d_2}^2 & 0 \\ 0 & 0 & {d_3}^2 \end{bmatrix} \qquad (3.10)$$

とすると

$$V=\boldsymbol{b}_1\boldsymbol{b}_1'+\boldsymbol{b}_2\boldsymbol{b}_2'+D=\begin{bmatrix}\boldsymbol{b}_1\boldsymbol{b}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{b}_1' \\ \boldsymbol{b}_2'\end{bmatrix}+D \qquad (3.11)$$

と表される。なお $\boldsymbol{b}_1'$、$\boldsymbol{b}_2'$ は $\boldsymbol{b}_1$、$\boldsymbol{b}_2$ を横ベクトルにしたものである。ここに出てきた行列 $[\boldsymbol{b}_1$、$\boldsymbol{b}_2]$ は (3.1) 式の係数行列そのものなので、因子負荷量行列と呼ばれる。以後これを B で表す。因

子分析のモデルの係数構造を推定する問題は、変量の個数を行数、想定する因子数を列数とする縦長の矩形行列 B を考えて、V を

$$V = BB' + D \tag{3.12}$$

と分解するような B と非負対角行列 D を求めることに帰着する。計算方法は次に示す。

3.3 因子負荷量の推定方法

因子分析で、変量 p 個、想定する因子数を m 個とすると、

$$V - D = BB' \tag{3.13}$$

(3.13) 式の V、D は $p \times p$ の対角行列で B、B' は $p \times m$、$m \times p$ の行列である。(3.13) 式を計算する。V から適当な対角行列を引くことにより、右辺が近似ではなく同じ $p \times m$ 行列とその転置行列の積になるような D と B を同時に決定することである。具体的な手順を下に示す。

1. 独自因子の分散の初期値 $D^{(0)}$ を設定する

2. k 回目の反復での推定値 $D^{(k)}$ を用いて、$V - D^{(k)}$ の固有値 $\lambda_1 \geq \lambda_2 \geq \cdots\cdots \geq \lambda_m$ と、対応する固有ベクトル $a_1, a_2, \cdots\cdots, a_m$ を計算し、
   $B^{(k)} = [\sqrt{\lambda_1}a_1\ \sqrt{\lambda_2}a_2\ \cdots\cdots\ \sqrt{\lambda_m}a_m]$
   とおく。

3. $V - D = B^{(k)}B^{(k)'}$ の対角成分を $\mathrm{D}^{(k+1)}$ とおく。

4. もし $\mathrm{D}^{(k)}$ と $\mathrm{D}^{(k+1)}$ との各成分が十分に近ければ終了する。そうでなければ $\mathrm{D}^{(k+1)}$ を新たな $\mathrm{D}^{(k)}$ として 2 に戻る。

これを繰り返して因子負荷量が求められる。このように因子分析の解を求めることを、主因子法とよぶ。

3.4 共通性と独自性

ここで、(3.7) 式を一般化して考えると、直交因子の場合、想定する因子数が m 個ならば、第 j 変量の分散は

$$Var(y_j) = {b_{j1}}^2 + {b_{j2}}^2 + \cdots\cdots + {b_{jm}}^2 + {d_j}^2 \tag{3.14}$$

となる。この値が行列 V の第 (j, j) 成分に等しくなることから、

$${b_{j1}}^2 + {b_{j2}}^2 + \cdots\cdots + {b_{jm}}^2 = v_{jj} - {d_j}^2 \tag{3.15}$$

という関係が成り立つ。左辺は、第 j 変量の m 個の共通因子に関する因子負荷量の 2 乗の和である。ここで相関行列の対角成分は 1 より

$${b_{j1}}^2 + {b_{j2}}^2 + \cdots\cdots + {b_{jm}}^2 = 1 - {d_j}^2 \tag{3.16}$$

と書ける。第 j 変量の m 個の共通因子に関する因子負荷量の２乗の和は 0 と 1 の間の値とわかる。これは、変量 y_j の全分散のうち共通因子で説明される割合を示しており、変量 y_j の共通性とよばれている。共通性を ${h_j}^2$ とかくと、

$$ {h_j}^2 = 1 - {d_j}^2 \tag{3.17} $$

と書ける。逆に ${d_j}^2$ のことは、変量 y_j の独自性とよぶ。これは変量 y_j の全分散のうち独自因子で説明される割合を示している。

共通性の初期値として 1 をもってくるのがもっとも簡単な方法なので、§3.3 の $D^{(0)}$ の初期値を 0 に設定すると考えやすくなる。

3.5 単純構造の発見方法

3.5.1 単純構造の説明

一般に有効性の高い分析結果を得るためには、無数ある解のうち、因子負荷量の構造が次に述べる単純構造になるものを選ぶとよい。

ここで、(3.1) 式を書き直して、絶対値の大きい係数を■、小さい係数を□で表す。因子軸を回転させた結果、つぎのような係数の構造になるとする。

$$ \left.\begin{aligned} y_{1i} &= \blacksquare\, f_{1i} + \square\, f_{2i} + e_{1i} \\ y_{2i} &= \blacksquare\, f_{1i} + \square\, f_{2i} + e_{2i} \\ y_{3i} &= \square\, f_{1i} + \blacksquare\, f_{2i} + e_{3i} \end{aligned}\right\} (i = 1, 2 \cdots\cdots n) \tag{3.18} $$

この場合、第 1 因子の負荷量の絶対値は第 1 変量と第 2 変量において大きく、第 3 変量においては小さくなっている。また第 2 因子は第 3 変量において大きく、第 1 変量と第 2 変量において小さくなっている。話を簡単にするため、絶対値の大きい係数をすべて正の値をとると仮定して説明する。第 1 因子得点が正値をとることは、第 1 変量と第 2 変量の観測値が大きく、第 1 因子得点が負値をとることは、第 1 変量と第 2 変量の観測値が小さくなる。第 3 変量には、第 1 因子得点はほとんど影響しない。また、第 2 因子得点が正値をとることは、第 3 変量の観測値が大きく、第 1 因子得点が負値をとることは、第 3 変量の観測値が小さくなる。第 1 変量と第 2 変量には第 2 因子得点はほとんど影響しない。

したがって第 1 因子はほぼ第 1 変量と第 2 変量に関連づけられる因子、第 2 因子はほぼ第 3 変量に関連づけられる因子となり、それぞれの変量の内容をみることにより、因子が何を意味しているか解釈が明確になる。

この例のように、ある特定因子の負荷量の絶対値はある一群の変量で大きく、他の変量では小さいとうい分離が生じたとき、そのような係数構造を単純構造とよぶ。

3.5.2 基準バリマックス法

計算によって、単純構造になるような最適な回転を得る方法にバリマックス法がある。

(3.1) 式の場合、回転後の因子負荷量行列ともとの因子負荷量行列の間の関係は、

$$\begin{bmatrix} {b_{11}}^* & {b_{12}}^* \\ {b_{21}}^* & {b_{22}}^* \\ {b_{31}}^* & {b_{32}}^* \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.19}$$

となる。左辺の回転後の因子負荷量行列を B^* とする。これを $p \times m$ で一般化すると、

$$\begin{bmatrix} {b_{11}}^* & \cdots\cdots & {b_{1m}}^* \\ \vdots & & \vdots \\ \vdots & & \vdots \\ {b_{p1}}^* & \cdots\cdots & {b_{pm}}^* \end{bmatrix} = \begin{bmatrix} b_{11} & \cdots\cdots & b_{1m} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ b_{p1} & \cdots\cdots & b_{pm} \end{bmatrix} \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & & \vdots \\ t_{m1} & \cdots & t_{mm} \end{bmatrix} \tag{3.20}$$

となる。回転後の因子負荷量行列 B^* の、第 k 列における因子負荷量の 2 乗 $({b_{jk}}^*)^2$ の分散は、

$$\begin{aligned} {\sigma_k}^2 &= \frac{1}{p}\sum_{j=1}^{p}\{({b_{jk}}^*)^2\}^2 - \frac{1}{p}\left\{\sum_{j=1}^{p}({b_{jk}}^*)^2\right\}^2 \\ &= \frac{1}{p^2}\left[p\sum_{j=1}^{p}({b_{jk}}^*)^4 - \left\{\sum_{j=1}^{p}\{({b_{jk}}^*)^2\right\}^2\right] \end{aligned} \tag{3.21}$$

となる。これを大きくすることが、個々の成分の 2 乗である $({b_{jk}}^*)^2$ を小さいものと大きいものの二つに分極させることに対応すると考えられる。さらに m 個の列全体についてこの量を集計して、

$$\sigma^2 = \sum_{k=1}^{m}{\sigma_k}^2 = \frac{1}{p^2}\sum_{k=1}^{m}\left[p\sum_{j=1}^{p}({b_{jk}}^*)^4 - \left\{\sum_{j=1}^{p}\{({b_{jk}}^*)^2\right\}^2\right] \tag{3.22}$$

という量を考えると、これを最大化することが単純構造につながると考えられる。この基準にもとずく回転を粗バリマックス法とよぶ。

ただし、共通性の大きい変量では因子負荷量の絶対値も平均的に大きいため、粗バリマックス法で計算すると、共通性の大きい変量の影響が強く出すぎる欠点がある。そこで、各変量の因子負荷量の 2 乗を共通性 (3.17) 式で割るという修正をほどこして、

$$S = \sum_{k=1}^{m}[p\sum_{j=1}^{p}(\frac{{b_{jk}}^*}{h_j})^4 - \{\sum_{j=1}^{p}\{(\frac{{b_{jk}}^*}{h_j})^2\}^2] \tag{3.23}$$

のような量を最大化するという基準がある。この基準にもとづく回転方法を基準バリマックス法とよぶ。

3.6 因子得点の推定方法

因子分析は因子負荷量が求まっても個々の変量のもつ因子得点はわからない。これは、因子分析が特定のモデルを想定した分析であり、実際の変量をモデルが記述する法則が偶然的なぶ

れを含みながら出る結果とみなしているからである。その意味では因子分析は、回帰分析に近い性質を備えている。このため、因子得点の推定は、最小二乗法によってモデルの理論値と実際の変量とのあいだの誤差を最小にすることが必要である。この段階で活用されるのは重回帰分析である。

先に変量 $p=3$ 因子数 $m=2$ のモデルで考えると、推定には、変量 f_1、f_2 を元の変量 y_1, y_2, y_3 の１次結合で表現する式を考え、その係数ベクトルを

$$\boldsymbol{c}_1 = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \end{bmatrix}, \boldsymbol{c}_2 = \begin{bmatrix} c_{12} \\ c_{22} \\ c_{32} \end{bmatrix} \tag{3.24}$$

とする。つまり、

$$\left. \begin{array}{l} f_1 = c_{11}y_1 + c_{21}y_2 + c_{31}y_3 \\ f_2 = c_{12}y_1 + c_{22}y_2 + c_{32}y_3 \end{array} \right\} \tag{3.25}$$

という重回帰式を、それぞれ最小二乗法で解く。ここにでてくる変量はすべて平均が 0 なので、定数項は最初から 0 と仮定することが許される。まず１番目の回帰式ついてだけ考えると c_{11}、c_{21}、c_{31} は、次の方程式の解として求められる。

$$\begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \end{bmatrix} = \begin{bmatrix} s_{1f} \\ s_{2f} \\ s_{3f} \end{bmatrix} \tag{3.26}$$

s_{jk} は y_i と y_k の偏差の積和、s_{jf} は y_i と f_1 の偏差の積和である。元の変量の個数 n で割ると、

$$\begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} \begin{bmatrix} c_{11} \\ c_{21} \\ c_{31} \end{bmatrix} = \begin{bmatrix} Cov(y_1, f_1) \\ Cov(y_2, f_1) \\ Cov(y_3, f_1) \end{bmatrix} \tag{3.27}$$

(3.27) 式の右辺の共分散ベクトルの各成分は、(3.1) 式と (3.7) 式を元に次のように変形できる。

$$\left. \begin{array}{l} Cov(y_1, f_1) = Cov(b_{11}f_1 + b_{12}f_2 + e_1, f_1) = b_{11}Var(f_1) = b_{11} \\ Cov(y_2, f_1) = Cov(b_{21}f_1 + b_{22}f_2 + e_2, f_1) = b_{21}Var(f_1) = b_{21} \\ Cov(y_3, f_1) = Cov(b_{31}f_1 + b_{32}f_2 + e_3, f_1) = b_{31}Var(f_1) = b_{31} \end{array} \right\} \tag{3.28}$$

つまり、(3.27) 式は、第１因子の因子負荷量ベクトルを $\boldsymbol{b}_1$ として、

$$V\boldsymbol{c}_1 = \boldsymbol{b}_1 \quad \therefore \boldsymbol{c}_1 = V^{-1}\boldsymbol{b}_1 \tag{3.29}$$

であり、因子負荷量ベクトルと相関行列から、因子得点の係数を求められる。同様にして、第２因子得点の係数は

$$\boldsymbol{c}_1 = V^{-1}\boldsymbol{b}_1 \tag{3.30}$$

で求まる。

因子分析の計算方法は、(3.8) 式と §(3.3) から因子負荷量を推定する。(3.23) 式で因子を回転させ、(3.30 式) によって因子得点を計算した。

4　分析の方法

この章では、実際の主成分分析の方法について説明する。過去４年分の中山競馬場２５００ｍの勝ち馬４２頭の右回り成績、左回り成績、競馬場の成績、距離成績と、過去４年分の有馬記念の出馬６１頭の右回り成績、左回り成績、競馬場の成績、距離成績、ＧＩの成績、重賞競争の成績からデータを組み合わせて３つを選び主成分分析と因子分析を行なった。

4.1　計算プログラムについて

まず、主成分分析をするにあたり、３要素を読み込み、標準化し、相関行列を求める。固有値・固有ベクトル、寄与率、累積寄与率をテキスト形式で返し、主成分得点をエクセル形式で返すプログラムを作成した。なお自分で計算した結果とプログラムが計算した結果を比べ、プログラムが正しいことを確認した。

次に、因子分析は、３要素を読み込み、標準化し、相関行列を求める。因子数は 2 つであると想定して計算する。バリマックス法で回転させた因子負荷量、因子得点の係数をテキスト形式で返し、因子得点をエクセル形式で返すプログラムを作成した。尚、§3 の (3.1) 式から因子負荷量と因子得点と独自因子を代入し、元の要素の値に戻るか確認しプログラムが正しいかを確認した。

4.2　計算方法

主成分分析の計算方法は、§2 で述べたように累積寄与率を決めるので、本研究での累積寄与率は約 80 %以上のものを採用とする。合成変量は３変量の場合次の式になる。

$$Z_i = a_1 y_{1i} + a_2 y_{2i} + a_3 y_{3i} \tag{4.1}$$

(4.1) 式から主成分得点と計算して散布図を作成する。主成分得点の固有ベクトルの値から、第１主成分得点と第２主成分得点に対してどの要素が強く働き、どのような説明ができるか読み取り、議論を進める。

因子分析の計算方法は、(3.30) 式から、因子得点を計算する。バリマックス法で回転させた、因子負荷量の値から、第１因子得点と第２因子得点がどの値で説明ができるかを調べる。

5　結果

この章では主成分分析と因子分析を行った結果を示す。分析のとき用いた要素は結果の部分で示すことにする。要素は、すべて標準化して議論する。ここで固有値は λ、固有ベクトルは a、因子負荷量は b、因子得点係数を c で表す。因子負荷量は回転後の値を示す。主成分分析の累積寄与率を約 80 %以上に定め、分析を行なう。なお表に記載している馬番は出走したときにスタートした馬の番号を示す。

5.1　中山競馬場 2500 m を用いての結果

ここでは、中山競馬場の 2500 m の結果から有馬記念の勝因を探る。用いた３変量の構成は、

1. ３着以内率
2. 距離の成績の１着に 10、２着に 5、３着に 2.5 を掛け、着外の回数を足した
3. 右回りの成績に１着に 5、２着に 2.5、３着に１を掛け、その中の中山競馬場の成績を２倍したもの

である。§5.1.1 では、中山競馬場の 2500 m の成績で主成分分析を行い、計算された固有ベクトルが勝因だと仮定し、過去有馬記念の出走馬と、2008 年有馬記念の出走馬に適用して、散布図を作成し、分析する。§5.1.2 では、中山競馬場の 2500 m の成績で因子分析を行い、計算された因子負荷量が勝因だと仮定し、因子得点係数を過去有馬記念の出走馬と、2008 年有馬記念の出走馬に適用して、散布図を作成し、分析する。§5.1.3 に結果の考察を書く。使用した元のデータを表 1 に示す。なお中山競馬場 2500 mの勝ち馬データは馬のクラスをわけて、2004 年 1 月から 2008 年 8 月までにのデータを記載し、過去有馬記念のデータは上から年代順に 1,2,3 着を示しており、それ以降は年代別の馬番の順に記載し、2008 年有馬記念は馬番順に記載している。この要素を標準化したものを用いた。

表 1: 中山競馬場 2500 mの分析に使用した元データ

中山競馬場2500mの勝ち馬の元データ				過去4年分の有馬記念の元データ					2008年有馬記念の元データ			
レースのクラス	3着率	距離の成績	中山競馬場と右回りの成績	年代	馬番	3着率	距離の成績	中山競馬場と右回りの成績	馬番	3着率	距離の成績	中山競馬場と右回りの成績
ＧＩレース	53.3	2.5	57.0	2007年	3番	53.3	2.5	57.0	1番	72.7	0.0	17.5
	100.0	5.0	60.0	2006年	4番	100.0	5.0	60.0	2番	30.8	0.0	7.5
	60.0	1.0	22.0	2005年	10番	60.0	1.0	22.0	3番	28.6	7.0	37.5
	85.7	7.5	38.0	2004年	1番	85.7	7.5	38.0	4番	56.3	12.0	46.0
GⅡレース	56.3	1.0	67.0	2007年	7番	100.0	0.0	30.0	5番	50.0	0.0	22.5
	44.4	0.0	40.0	2006年	1番	72.2	35.0	26.0	6番	64.3	0.0	45.5
	60.0	7.5	41.5	2005年	6番	100.0	0.0	40.0	7番	69.2	7.5	15.5
	26.9	3.0	62.0	2004年	9番	65.7	30.5	54.5	8番	70.6	10.0	27.0
	60.0	3.5	27.5	2007年	4番	61.5	2.5	56.5	9番	72.0	2.0	53.5
1600万以下のレース	33.3	2.0	6.0	2006年	5番	60.0	0.0	49.5	10番	60.0	22.5	89.5
	52.2	2.5	32.0	2005年	14番	60.0	5.0	39.5	11番	53.3	0.0	37.0
	50.0	25.0	11.0	2004年	6番	62.5	0.0	48.0	12番	53.3	0.0	14.5
	33.3	10.0	16.0	2007年	1番	80.0	1.0	48.5	13番	100.0	5.0	40.0
	36.7	32.5	30.5	2007年	2番	86.7	1.0	30.5	14番	52.3	20.5	69.0
	77.3	0.0	30.0	2007年	5番	72.7	0.0	14.5				
	57.1	0.0	20.0	2007年	6番	76.0	50.0	37.0				
	37.5	10.0	19.0	2007年	8番	100.0	0.0	21.0				
	44.4	0.0	15.0	2007年	9番	41.7	0.0	20.0				
	31.3	1.0	10.0	2007年	10番	55.0	1.0	35.5				
	58.3	0.0	17.5	2007年	11番	32.0	5.0	37.5				
1000万以下のレース	47.4	0.0	29.5	2007年	12番	55.6	10.0	28.5				
	31.8	3.5	10.5	2007年	13番	36.0	15.0	36.0				
	44.4	1.0	10.0	2007年	14番	45.5	1.0	9.0				
	41.7	0.0	7.0	2007年	15番	55.0	1.0	37.5				
	27.3	0.0	6.0	2007年	16番	80.0	0.0	26.0				
	34.5	1.0	20.5	2006年	2番	47.4	14.0	36.0				
	40.0	26.0	15.5	2006年	3番	100.0	0.0	28.0				
	31.3	4.5	17.5	2006年	6番	63.2	0.0	43.5				
	100.0	0.0	10.0	2006年	7番	40.0	4.0	37.5				
	46.7	12.0	10.0	2006年	8番	78.6	0.0	36.0				
	37.5	2.5	18.0	2006年	9番	37.0	37.0	39.5				
	30.4	17.0	23.5	2006年	10番	75.0	0.0	22.0				
	35.7	0.0	9.0	2006年	11番	57.7	22.0	30.0				
	25.8	15.5	27.0	2006年	12番	66.7	0.0	29.0				
	60.0	0.0	17.0	2006年	13番	44.4	21.0	39.5				
	40.9	0.0	6.0	2006年	14番	40.0	0.0	11.0				
	28.6	1.0	10.0	2005年	1番	30.6	1.0	32.5				
	26.8	0.0	30.5	2005年	2番	47.8	0.0	34.5				
	25.0	0.0	5.0	2005年	3番	88.9	17.5	49.0				
	30.0	0.0	20.0	2005年	4番	46.2	2.0	30.0				
	50.0	2.0	10.0	2005年	5番	47.1	23.0	28.5				
	50.0	1.0	20.0	2005年	7番	59.4	0.0	50.5				
				2005年	8番	48.6	0.0	58.0				
				2005年	9番	62.5	35.5	59.5				
				2005年	11番	72.7	22.5	23.0				
				2005年	12番	33.3	11.0	38.0				
				2005年	13番	64.0	13.5	42.0				
				2005年	15番	53.3	13.0	35.0				
				2005年	16番	60.0	0.0	34.5				
				2004年	2番	77.8	0.0	27.0				
				2004年	3番	60.0	0.0	1.0				
				2004年	4番	75.0	0.0	30.0				
				2004年	5番	66.7	0.0	17.0				
				2004年	7番	22.9	2.0	58.0				
				2004年	8番	63.6	3.0	69.0				
				2004年	10番	58.3	11.0	25.0				
				2004年	11番	48.0	1.0	32.0				
				2004年	12番	36.4	0.0	12.5				
				2004年	13番	51.6	1.0	28.0				
				2004年	14番	76.5	10.0	38.5				
				2004年	15番	64.3	0.0	27.0				

5.1.1 中山競馬場 2500 m の主成分分析の結果

(2.10) 式で計算した、固有値と固有ベクトルは、固有値の大きい順番に、

$$\lambda = 1.343807, 1.016981, 0.639212$$

であり、固有ベクトルは

$$a = \begin{pmatrix} 0.714094 \\ -0.181165 \\ 0.676202 \end{pmatrix}, \begin{pmatrix} -0.062886 \\ 0.945428 \\ 0.319704 \end{pmatrix}, \begin{pmatrix} -0.697220 \\ -0.270822 \\ 0.663732 \end{pmatrix}$$

となった。固有値は 2 つで寄与率が約 79 %で累積寄与率が約 80 %なので、第 2 主成分までで議論してよいことが確認された。

上記の固有値と固有ベクトルを中山競馬場の 2500 mでの勝因だと仮定し、過去有馬記念の出走馬と、2008 年有馬記念の出走馬の同じ要素に上記の固有ベクトルを用いて、(4.1) 式より主成分得点を計算した。計算された主成分得点を表 2 に載せる。表のデータは、右から過去有馬記念の出走馬の第 1 主成分得点、第 2 主成分得点、2008 年有馬記念の出走馬の第 1 主成分得点、第 2 主成分得点である。また過去有馬記念の上から年代順に 1,2,3 着を示しており、それ以降は年代別の馬番の順に記載し、2008 年有馬記念は馬番順に記載している。

表 2: 中山競馬場の 2500 mから分析した過去有馬記念と 2008 年有馬記念の主成分得点

過去有馬記念出走馬				2008年有馬記念出走馬		
年代	馬番	第1主成分得点	第2主成分得点	馬番	第1主成分得点	第2主成分得点
2007年	3番	0.846871	0.151525	1番	0.081972	−1.123965
2006年	4番	2.739921	0.273892	2番	−1.94971	−1.11786
2005年	10番	−0.609463	−0.817442	3番	−1.286811	0.219337
2004年	1番	1.062047	0.015116	4番	−0.007583	0.880101
2007年	7番	1.331753	−0.848038	5番	−0.697735	−0.968896
2006年	1番	−0.489898	2.079517	6番	0.599174	−0.68564
2005年	6番	1.827858	−0.613482	7番	−0.306801	−0.184499
2004年	9番	0.747787	2.393434	8番	0.044068	0.296599
2007年	4番	1.135368	0.112206	9番	1.113564	−0.342078
2006年	5番	0.770857	−0.256067	10番	1.228494	2.838687
2005年	14番	0.194603	−0.07236	11番	−0.115215	−0.769618
2004年	6番	0.79196	−0.299662	12番	−0.808486	−1.097392
2007年	1番	1.469367	−0.263164	13番	1.776115	−0.257804
2007年	2番	0.832368	−0.707907	14番	0.328954	2.313028
2007年	5番	−0.480277	−1.119743			
2007年	6番	−0.039438	3.57953			
2007年	8番	0.885258	−1.059138			
2007年	9番	−1.391852	−0.886432			
2007年	10番	−0.130758	−0.483969			
2007年	11番	−0.974429	−0.02506			
2007年	12番	−0.599375	0.102696			
2007年	13番	−1.056311	0.762824			
2007年	14番	−1.808409	−1.073576			
2007年	15番	−0.031537	−0.437058			
2007年	16番	0.36916	−0.874566			
2006年	2番	−0.604716	0.640814			
2006年	3番	1.232531	−0.894949			
2006年	6番	0.595458	−0.407567			
2006年	7番	−0.652739	−0.13563			
2006年	8番	0.811775	−0.6353			
2006年	9番	−1.19712	2.681909			
2006年	10番	−0.02032	−0.951565			
2006年	11番	−0.637079	1.134644			
2006年	12番	0.009831	−0.759449			
2006年	13番	−0.657909	1.31857			
2006年	14番	−1.9033	−1.091812			
2005年	1番	−1.211854	−0.472237			
2005年	2番	−0.439433	−0.566851			
2005年	3番	1.56973	1.098885			
2005年	4番	−0.755872	−0.499712			
2005年	5番	−1.132524	1.218779			
2005年	7番	0.797543	−0.230592			
2005年	8番	0.756981	−0.018337			
2005年	9番	0.793428	2.939742			
2005年	11番	−0.419255	0.961811			
2005年	12番	−0.996132	0.484209			
2005年	13番	0.335207	0.683867			
2005年	15番	−0.412872	0.513854			
2005年	16番	0.026699	−0.6079			
2004年	2番	0.334714	−0.843708			
2004年	3番	−1.635255	−1.393661			
2004年	4番	0.376564	−0.763921			
2004年	5番	−0.585495	−1.040916			
2004年	7番	−0.257012	0.23544			
2004年	8番	1.827721	0.440161			
2004年	10番	−0.685881	0.09517			
2004年	11番	−0.571848	−0.54251			
2004年	12番	−1.966431	−1.044516			
2004年	13番	−0.632743	−0.648445			
2004年	14番	0.695268	0.26693			
2004年	15番	−0.181088	−0.798285			

次に散布図を図 1 に載せる。図の x 軸は第 1 主成分得点、y 軸は第 2 主成分得点、点の◆は過去有馬記念の出走馬の主成分得点、■は過去の 1,2,3 着の主成分得点、▲は 2008 年有馬記念の出走馬の主成分得点である。

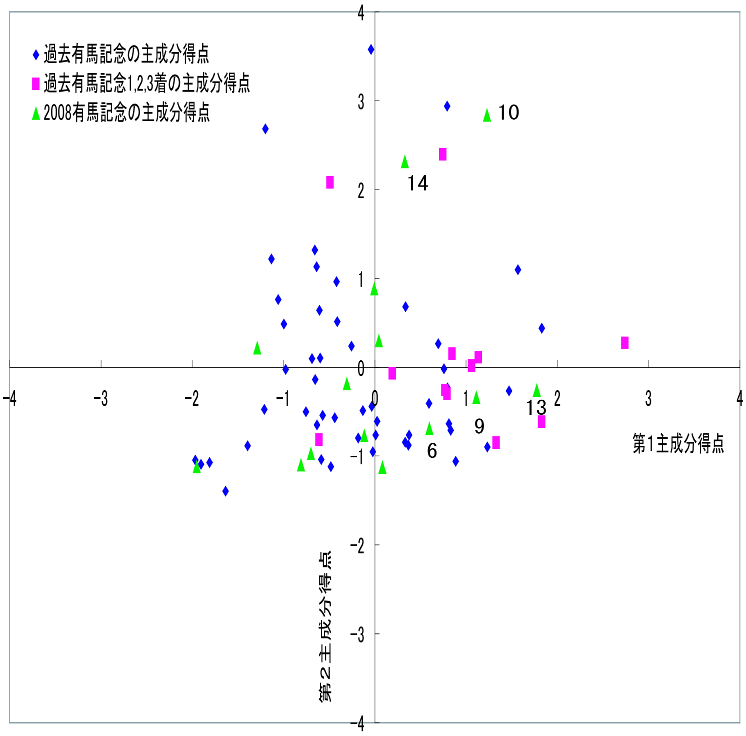

図 1: 中山競馬場の 2500 mから分析した過去有馬記念と 2008 年有馬記念の主成分得点の散布図

図 1 より、第 1 主成分得点が正の方向に相関が見られる。第 1 主成分得点が表す値に注目すると、正の値は 3 着以内率が高く右回りの中山競馬場が得意な馬を表していることがわかる。
第 2 主成分得点が表す値に注目すると、正の値は距離が得意な馬を表していることがわかる。
2 主成分が表す値から、3 着以内率が高く右回りの中山競馬場が得意な馬が勝ち易いことがわかる。図 1 より 2008 年有馬記念の出走馬の第 1 主成分得点が正の値をとっている馬は 6, 9, 10, 13, 14 番である。

5.1.2 中山競馬場 2500 m の因子分析の結果

(3.13) 式を因子負荷量の計算方法を用いて計算し、(3.23) 式によって回転された因子負荷量は

$$b = \begin{pmatrix} 0.410371 \\ -0.012447 \\ 0.984753 \end{pmatrix}, \begin{pmatrix} -0.132816 \\ 0.998062 \\ 0.050199 \end{pmatrix}$$

である。また (3.30) 式で計算した因子得点係数は、

$$c = \begin{pmatrix} 0.089109 \\ -0.028962 \\ 0.955919 \end{pmatrix}, \begin{pmatrix} -0.030823 \\ 0.993647 \\ 0.032761 \end{pmatrix}$$

となる。上記の因子負荷量が勝因だと仮定し、因子得点係数を過去有馬記念の出走馬と、2008 年有馬記念の出走馬の同じ要素に適用し、計算されたもの因子得点を表 3 に載せる。表のデータは、右から過去有馬記念の出走馬の第 1 因子得点、第 2 因子得点、2008 年有馬記念の出走馬の第 1 因子得点、第 2 因子得点である。また過去有馬記念の上から年代順に 1,2,3 着を示しており、それ以降は年代別の馬番の順に記載し、2008 年有馬記念は馬番順に記載している。

表 3: 中山競馬場の 2500 mから分析した過去有馬記念と 2008 年有馬記念の因子得点

過去有馬記念出走馬				2008年有馬記念出走馬		
年代	馬番	第1因子得点	第2因子得点	馬番	第1因子得点	第2因子得点
2007年	3番	1.51556	−0.34621	1番	−0.771642	−0.880638
2006年	4番	1.942206	−0.196217	2番	−1.422296	−0.821172
2005年	10番	−0.903286	−0.573262	3番	−0.154188	0.165217
2004年	1番	0.324707	−0.005715	4番	0.338717	0.798508
2007年	7番	−0.148953	−0.707919	5番	−0.670375	−0.83287
2006年	1番	−0.651718	2.405477	6番	0.404856	−0.823925
2005年	6番	0.552371	−0.683883	7番	−0.906007	0.127268
2004年	9番	1.327597	2.089063	8番	−0.407668	0.476841
2007年	4番	1.519589	−0.360934	9番	0.785034	−0.55774
2006年	5番	1.027917	−0.595083	10番	2.211472	2.263423
2005年	14番	0.313781	−0.179524	11番	−0.021849	−0.817083
2004年	6番	0.934638	−0.602811	12番	−1.001897	−0.850671
2007年	1番	1.050578	−0.542551	13番	0.329018	−0.225737
2007年	2番	−0.179861	−0.596864	14番	1.286824	1.978579
2007年	5番	−1.366165	−0.700152			
2007年	6番	0.099417	3.744434			
2007年	8番	−0.780145	−0.729551			
2007年	9番	−1.128238	−0.635809			
2007年	10番	0.019663	−0.532568			
2007年	11番	0.040018	−0.138155			
2007年	12番	−0.491467	0.240888			
2007年	13番	−0.071735	0.730833			
2007年	14番	−1.884139	−0.580596			
2007年	15番	0.159927	−0.527761			
2007年	16番	−0.524838	−0.68455			
2006年	2番	−0.01482	0.624114			
2006年	3番	−0.289218	−0.712726			
2006年	6番	0.62238	−0.614782			
2006年	7番	0.080723	−0.239267			
2006年	8番	0.169811	−0.658206			
2006年	9番	0.122119	2.671814			
2006年	10番	−0.829207	−0.685918			
2006年	11番	−0.407007	1.296058			
2006年	12番	−0.377852	−0.655405			
2006年	13番	0.198402	1.252906			
2006年	14番	−1.767535	−0.654637			
2005年	1番	−0.307068	−0.499539			
2005年	2番	−0.082235	−0.611017			
2005年	3番	1.085795	0.894636			
2005年	4番	−0.410585	−0.443356			
2005年	5番	−0.565307	1.397852			
2005年	7番	1.095189	−0.59169			
2005年	8番	1.56969	−0.555852			
2005年	9番	1.650189	2.545953			
2005年	11番	−0.827698	1.298455			
2005年	12番	0.065907	0.388417			
2005年	13番	0.486401	0.5672			
2005年	15番	−0.05426	0.524061			
2005年	16番	−0.024068	−0.631137			
2004年	2番	−0.465195	−0.678518			
2004年	3番	−2.373503	−0.711656			
2004年	4番	−0.268147	−0.66669			
2004年	5番	−1.219441	−0.684248			
2004年	7番	1.442033	−0.337631			
2004年	8番	2.404975	−0.290394			
2004年	10番	−0.72662	0.315942			
2004年	11番	−0.259175	−0.529437			
2004年	12番	−1.6795	−0.645095			
2004年	13番	−0.522541	−0.544988			
2004年	14番	0.309504	0.230456			
2004年	15番	−0.52956	−0.656255			

次に散布図を図 2 に載せる。図の x 軸は第 1 因子得点、y 軸は第 2 因子得点、点の◆は過去有馬記念の出走馬の因子得点、■は過去の 1,2,3 着の因子得点、▲は 2008 年有馬記念の出走馬の因子得点である。

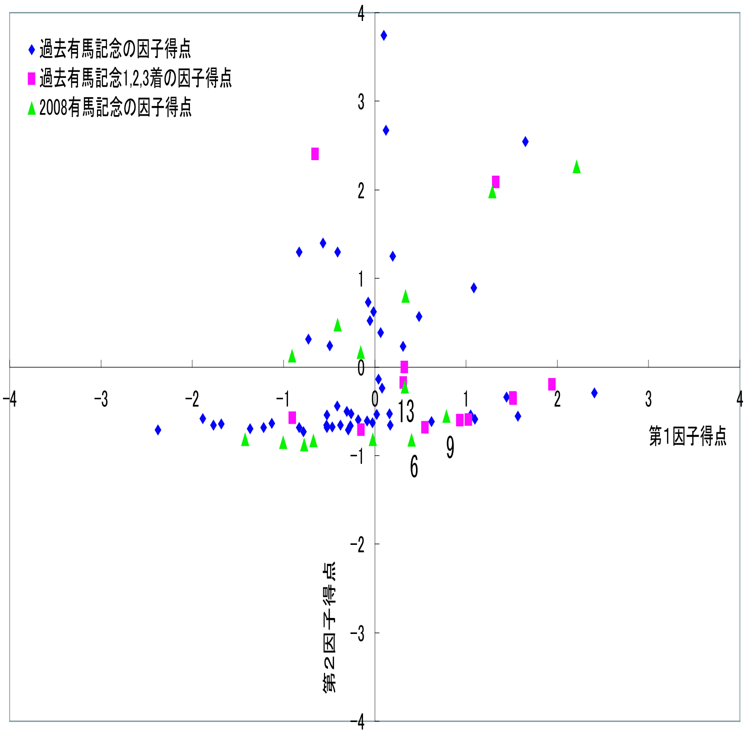

図 2: 中山競馬場の 2500 mから分析した過去有馬記念と 2008 年有馬記念の因子得点の散布図

図 2 より、第 1 因子得点が正の方向、第 2 因子得点が負の方向に相関が見られる。

第 1 因子得点が表す値に注目すると、正の値は右回りの中山競馬場が得意な馬を表していることがわかる。

第 2 因子得点が表す値に注目すると、正の値は距離が得意な馬を表していることがわかる。

２因子が表す値から、右回りの中山競馬場が得意な馬で、３着以内率を高く持っている馬が勝ち易いことがわかる。図 2 より 2008 年有馬記念の出走馬の第 1 因子得点が正の値をとり、第 2 因子得点が負の値をとる馬は 6, 9, 13 番である。

5.1.3　中山競馬場 2500 m からの結果の考察

以上より、主成分分析より、6, 9, 10, 13, 14 番が勝ち易いといえ、因子分析より 6, 9, 13 番が勝ち易いといえる。特に、どちらの分析にも出ている 6, 9, 13 番は勝つ確率が高いといえる。

主成分分析の第 1 主成分得点と、因子分析の第 1 因子得点を比べると、第 1 主成分得点が正の値は 3 着以内率が高く右回りの中山競馬場が得意な馬を表すが、第 1 因子得点が正の値は右回りの中山競馬場が得意な馬だけで表せる値に変わっている。ただ因子分析の因子負荷量の結果も 3 着以内率を正の値をとっているため、お互いの得点を示す値は大きく変わらないといえる。

次に主成分分析の第 2 主成分得点と、因子分析の第 2 因子得点を比べると、お互い同じ距離が得意な馬が正の値を表している。

以上のことから主成分分析も因子分析も値は同じような事を示しているが、固有ベクトルと因子負荷量の値が大きく違うため、グラフは異なったものになり、相関の見える馬が減り、勝つ確率の高い馬の数を減らすことができた。

5.2 過去有馬記念を用いての結果１

ここでは、過去有馬記念の結果から有馬記念の勝因を探る。用いた３変量の構成は、

1. 距離の成績を主成分分析し、その時得られた第１主成分得点
2. 右回りと中山の成績を足したものを主成分分析し、その時得られた第１主成分得点
3. 重賞とＧＩの成績を足したものを主成分分析し、その時得られた第１主成分得点

である。§5.2.1 では、過去有馬記念の出走馬の成績で主成分分析を行い、計算された固有ベクトルが勝因だと仮定し、2008 年有馬記念の出走馬に適用して、散布図を作成し、分析する。§5.2.2 では、過去有馬記念の成績の出走馬の成績で因子分析を行い、計算された因子負荷量が勝因だと仮定し、因子得点係数を、2008 年有馬記念の出走馬に適用して、散布図を作成し、分析する。§5.2.3 に結果の考察を書く。使用した元のデータを表 4 に示す。過去有馬記念のデータは上から年代順に 1,2,3 着を示しており、それ以降は年代別の馬番の順に記載し、2008 年有馬記念は馬番順に記載している。要素は第１主成分得点である。

表 4: 過去有馬記念からの分析に使用した元データ 1

過去有馬記念の元データ			2008年有馬記念の元データ		
第１主成分得点					
距離の成績	右中山の成績	重賞成績	距離の成績	右中山の成績	重賞成績
−0.727234	0.000476	−1.22742	−0.984258	−1.395483	0.066648
−0.007601	−0.568275	1.084327	−0.984258	−1.155058	−1.59
−0.727234	−0.581521	0.943041	1.603477	−0.214598	0.207051
0.274951	0.50559	1.144161	0.525051	−0.409202	−1.64335
−1.009785	−1.06713	−0.16489	−0.984258	−0.477111	−0.75125
2.288531	0.210531	−1.68144	−0.984258	1.180212	0.611759
−1.009785	−1.412582	−0.58838	0.600531	−0.591115	−0.45324
4.038047	2.784839	2.068311	−0.214302	−0.046494	−0.71876
−0.727234	1.378404	2.543148	−0.244906	0.345205	3.294113
−1.009785	0.643367	1.024926	0.92533	0.839592	−0.04005
−0.007601	0.477083	0.527605	−0.984258	−0.702353	−0.35334
−1.009785	0.671875	−0.34991	−0.984258	−0.888406	0.225475
−0.727234	−0.194316	1.768769	0.230854	−0.378563	1.617819
−0.727234	1.148481	1.397064	2.479516	3.893373	−0.47288
−1.009785	−0.962103	−1.83759			
4.201787	1.726236	0.809373			
−1.009785	−1.419204	−0.92955			
−1.009785	−0.721678	−1.1018			
−0.727234	0.208151	1.569319			
0.402974	−0.174811	0.079962			
−0.098714	−1.038622	0.096066			
1.314045	−0.65804	−0.47552			
−0.727234	−0.690791	−0.3338			
−0.727234	−0.686548	−1.68144			
−1.009785	−0.824325	0.11855			
1.031493	−0.65804	−0.47552			
−1.009785	0.851042	1.107244			
−1.009785	−0.789195	1.604565			
0.120422	−0.174811	0.079962			
−1.009785	−0.65804	−0.18737			
3.290716	−0.034654	−0.80393			
−1.009785	−0.581521	−0.34991			
2.470759	−0.555393	−0.81198			
−1.009785	0.153516	−1.0857			
1.094908	−0.034654	−1.09375			
−1.009785	−1.585489	−1.68144			
−0.727234	−0.76969	−1.05683			
−1.009785	−0.117796	−0.19375			
1.186022	0.997822	3.51741			
−0.444682	−1.06713	−0.20986			
2.753311	−0.526886	−0.7895			
−1.009785	1.481051	−0.92317			
−1.009785	2.37337	−0.51411			
5.040232	3.379718	2.804103			
1.094908	−0.76731	−0.78783			
0.183837	1.017327	0.371453			
1.122607	2.832851	1.696606			
0.748941	−0.983987	−0.77339			
−1.009785	−0.117796	0.244166			
−1.009785	−0.498378	−1.83759			
−1.009785	−1.751774	−0.78783			
−1.009785	−1.06713	−0.20986			
−1.009785	−1.1764	−0.50606			
−0.444682	0.326423	−1.53972			
−0.16213	3.388721	0.41004			
0.183837	−1.150272	−0.92955			
−0.727234	−0.415236	−0.61086			
−1.009785	−1.613997	−1.39162			
−0.727234	0.339305	1.403445			
0.557503	2.209465	0.811044			
−1.009785	−1.010115	0.698189			

5.2.1 過去有馬記念の主成分分析の結果１

(2.10) 式で計算した、固有値と固有ベクトルは、固有値の大きい順番に、

$$\lambda = 1.823458, 0.770019, 0.406522$$

であり、固有ベクトルは

$$a = \begin{pmatrix} 0.504238 \\ 0.645159 \\ 0.574034 \end{pmatrix}, \begin{pmatrix} -0.800989 \\ 0.100985 \\ 0.590100 \end{pmatrix}, \begin{pmatrix} 0.322739 \\ -0.757346 \\ 0.567686 \end{pmatrix}$$

となる。固有値は 2 つで約 86 %なので 80 %を超えたので、第 2 主成分までで議論してよいことが確認された。

上記の固有値と固有ベクトルを勝因だと仮定し、2008 年有馬記念の出走馬の同じ要素に上記の固有ベクトルを掛け、(4.1) 式より主成分得点を計算した。計算された主成分得点を表 5 に載せる。表のデータは、右から過去有馬記念の出走馬の第 1 主成分得点、第 2 主成分得点、2008 年有馬記念の出走馬の第 1 主成分得点、第 2 主成分得点である。また過去有馬記念の上から年代順に 1,2,3 着を示しており、それ以降は年代別の馬番の順に記載し、2008 年有馬記念は馬番順に記載している。

表 5: 過去有馬記念からの分析の過去有馬記念と 2008 年有馬記念の主成分得点 1

過去有馬記念出走馬				2008年有馬記念出走馬		
年代	馬番	第1主成分得点	第2主成分得点	馬番	第1主成分得点	第2主成分得点
2007年	3番	−0.837252	−0.200826	1番	−1.134544	0.655286
2006年	4番	0.215392	0.488864	2番	−1.793274	−0.127629
2005年	10番	−0.10858	0.815077	3番	0.741199	−1.111316
2004年	1番	0.906907	0.450924	4番	−0.7341	−1.218961
2007年	7番	−0.990336	0.387745	5番	−1.056155	0.331934
2006年	1番	0.102941	−2.067901	6番	0.422125	1.122634
2005年	6番	−1.373998	0.151728	7番	−0.230119	−0.713484
2004年	9番	3.852075	−0.982069	8番	−0.462418	−0.191938
2007年	4番	1.685479	1.760756	9番	1.610489	1.804257
2006年	5番	0.476871	1.111782	10番	0.838865	−0.642426
2005年	14番	0.499921	0.302078	11番	−0.982474	0.506747
2004年	6番	−0.163645	0.440268	12番	−0.803828	0.772227
2007年	1番	0.488836	1.251687	13番	0.678993	0.58122
2007年	2番	1.018066	1.180079	14番	2.905242	−1.768549
2007年	5番	−1.732619	−0.423467			
2007年	6番	2.751678	−1.776532			
2007年	8番	−1.540151	−0.016039			
2007年	9番	−1.255307	−0.043026			
2007年	10番	0.605481	1.187072			
2007年	11番	0.085946	−0.197262			
2007年	12番	−0.534868	0.015951			
2007年	13番	−0.117091	−1.011368			
2007年	14番	−0.774862	0.180245			
2007年	15番	−1.415178	−0.479938			
2007年	16番	−0.727451	0.546665			
2006年	2番	−0.215107	−0.855667			
2006年	3番	0.62538	1.169232			
2006年	6番	−0.000442	1.277915			
2006年	7番	−0.01207	−0.041562			
2006年	8番	−0.78583	0.410413			
2006年	9番	0.739995	−2.210242			
2006年	10番	−0.823068	0.33705			
2006年	11番	0.17775	−1.805235			
2006年	12番	−0.787182	0.036939			
2006年	13番	−0.159908	−1.142294			
2006年	14番	−1.986135	−0.398267			
2005年	1番	−1.161109	−0.180639			
2005年	2番	−0.504646	0.451775			
2005年	3番	2.613478	1.152641			
2005年	4番	−0.815744	0.054304			
2005年	5番	0.301485	−1.947567			
2005年	7番	−0.011259	0.225927			
2005年	8番	0.653234	0.499904			
2005年	9番	4.863523	−1.124693			
2005年	11番	−0.399502	−1.052684			
2005年	12番	0.776104	0.164539			
2005年	13番	2.688752	0.446251			
2005年	15番	−0.626631	−0.872808			
2005年	16番	−0.295847	0.666418			
2004年	2番	−1.488649	−0.385279			
2004年	3番	−1.647547	0.026037			
2004年	4番	−1.011776	0.365704			
2004年	5番	−1.210492	0.211525			
2004年	7番	−0.716658	−0.482757			
2004年	8番	1.922099	0.569383			
2004年	10番	−0.9846	−0.651638			
2004年	11番	−0.761991	0.06714			
2004年	12番	−1.862949	−0.258562			
2004年	13番	0.595394	1.116571			
2004年	14番	1.742518	0.272265			
2004年	15番	−0.548827	0.81547			

次に散布図を図 3 に載せる。図の x 軸は第 1 主成分得点、y 軸は第 2 主成分得点、点の◆は過去有馬記念の出走馬の主成分得点、■は過去の 1,2,3 着の主成分得点、▲は 2008 年の出走馬の主成分得点である。

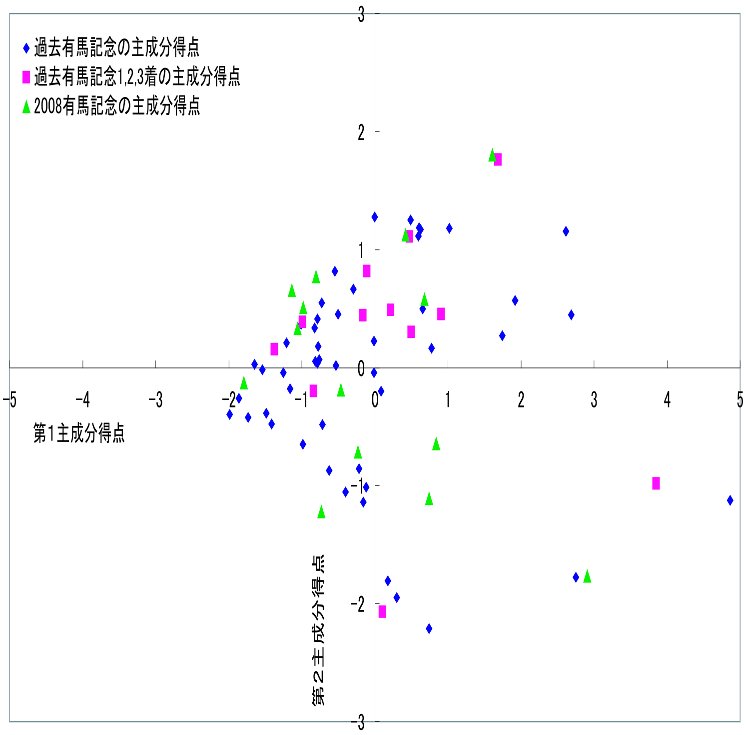

図 3: 過去有馬記念からの主成分得点の散布図 1

図 3 より第 2 主成分得点が正の値に相関が見られる。
第 1 主成分得点が表す値は、固有値から正の値はすべての第 1 主成分得点が大きい馬を表していることがわかる。
第 2 主成分得点が表す値は、固有値から正の値は特に重賞の成績の第 1 主成分得点が大きい馬を表していることがわかる。
２主成分から重賞の成績の第 1 主成分得点が大きい馬が勝ち易いことがわかる。図 3 より第２主成分得点が正の値をとる馬が多く有意な予想に用いるには適さないと判断する。

5.2.2 過去有馬記念を用いての因子分析の結果 1

(3.13) 式を因子負荷量の計算方法を用いて計算し、(3.23) 式によって回転された因子負荷量は

$$b = \begin{pmatrix} 0.120464 \\ 0.712798 \\ 0.934867 \end{pmatrix}, \begin{pmatrix} -0.974500 \\ -0.442229 \\ -0.066839 \end{pmatrix}$$

である。(3.30) 式で計算した因子得点係数は、

$$c = \begin{pmatrix} -0.232375 \\ 0.387769 \\ 0.778364 \end{pmatrix}, \begin{pmatrix} -0.963781 \\ -0.164373 \\ 0.252665 \end{pmatrix}$$

となる。上記の因子負荷量が勝因だと仮定し、因子得点係数を 2008 年有馬記念の要素に同じ因子得点係数を掛け、因子得点を計算する。表 6 に過去有馬記念の出走馬と 2008 年有馬記念の因子得点を載せる。表のデータは、右から過去有馬記念の出走馬の第 1 因子得点、第 2 因子得点、2008 年有馬記念の出走馬の第 1 因子得点、第 2 因子得点である。また過去有馬記念の上から年代順に 1,2,3 着を示しており、それ以降は年代別の馬番の順に記載し、2008 年有馬記念は馬番順に記載している。

表 6: 過去有馬記念からの因子得点 1

過去有馬記念出走馬				2008年有馬記念出走馬		
年代	馬番	第1因子得点	第2因子得点	馬番	第1因子得点	第2因子得点
2007年	3番	−0.677133	0.224533	1番	−0.168045	1.075558
2006年	4番	0.522551	0.308775	2番	−1.152859	0.701286
2005年	10番	0.542064	0.758048	3番	−0.279462	−1.366863
2004年	1番	0.855637	−0.009956	4番	−1.286871	−0.75864
2007年	7番	−0.282616	0.777968	5番	−0.411665	0.787948
2006年	1番	−1.386359	−1.898493	6番	0.961353	0.857326
2005年	6番	−0.665644	0.735398	7番	−0.598496	−0.556246
2004年	9番	1.572258	−2.616624	8番	−0.426705	0.049112
2007年	4番	2.196314	0.831142	9番	2.260898	0.85788
2006年	5番	1.027501	0.79839	10番	0.029131	−0.945739
2005年	14番	0.49318	0.051817	11番	−0.225937	0.899382
2004年	6番	0.147667	0.50604	12番	0.087188	1.043271
2007年	1番	1.198349	0.879438	13番	0.868221	0.177048
2007年	2番	1.382648	0.621439	14番	0.34325	−2.821323
2007年	5番	−1.330832	0.412849			
2007年	6番	0.397414	−2.847501			
2007年	8番	−0.888309	0.664686			
2007年	9番	−0.779106	0.535038			
2007年	10番	1.196667	0.783633			
2007年	11番	−0.068003	−0.226976			
2007年	12番	−0.250539	0.224832			
2007年	13番	−0.725584	−0.882861			
2007年	14番	−0.317988	0.50473			
2007年	15番	−1.187913	0.22134			
2007年	16番	−0.02259	0.804906			
2006年	2番	−0.680414	−0.695516			
2006年	3番	1.146391	0.787829			
2006年	6番	0.949249	1.11206			
2006年	7番	−0.022833	−0.039632			
2006年	8番	−0.167792	0.718414			
2006年	9番	−1.056781	−2.345963			
2006年	10番	−0.248676	0.674048			
2006年	11番	−1.095569	−1.734184			
2006年	12番	−0.491946	0.421105			
2006年	13番	−0.89312	−0.950867			
2006年	14番	−1.427001	0.529179			
2005年	1番	−0.810386	0.363567			
2005年	2番	−0.001084	0.64466			
2005年	3番	2.39998	−0.181954			
2005年	4番	−0.402029	0.393843			
2005年	5番	−1.117189	−1.920631			
2005年	7番	0.03292	0.27727			
2005年	8番	0.579545	0.243508			
2005年	9番	2.075854	−3.206439			
2005年	11番	−0.927013	−0.788458			
2005年	12番	0.532455	−0.180302			
2005年	13番	1.813205	−0.768001			
2005年	15番	−0.930891	−0.526994			
2005年	16番	0.282037	0.736564			
2004年	2番	−1.184195	0.35069			
2004年	3番	−0.901849	0.739007			
2004年	4番	−0.311689	0.768531			
2004年	5番	−0.53774	0.721016			
2004年	7番	−0.821143	−0.072044			
2004年	8番	1.36258	−0.260678			
2004年	10番	−0.994087	−0.162786			
2004年	11番	−0.409975	0.409649			
2004年	12番	−1.248644	0.593823			
2004年	13番	1.1309	0.731241			
2004年	14番	1.133892	−0.495601			
2004年	15番	0.293405	0.951456			

次に散布図を図 4 に載せる。図の x 軸は第 1 因子得点、y 軸は第 2 因子得点、点の◆は過去有馬記念の出走馬の因子得点、■は過去の 1,2,3 着の因子得点、▲は 2008 年の出走馬の主成分得点である。

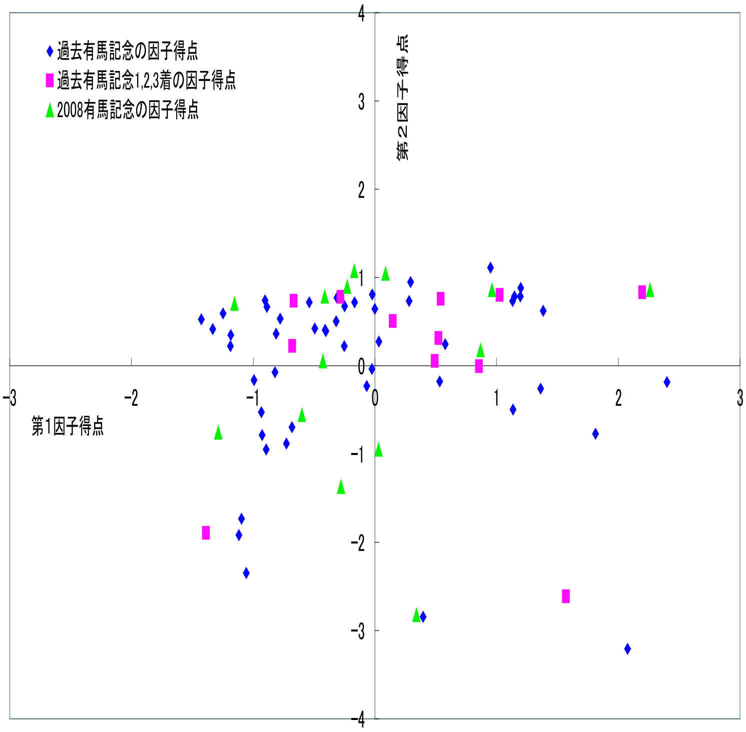

図 4: 過去有馬記念からの因子得点の散布図 1

図 4 より第 2 因子得点が正の値に相関が見られる。

第 1 因子得点が表す値は、因子負荷量から正の値は重賞の成績の第 1 主成分得点と右回りの第 1 主成分得点が大きい馬を表していることがわかる。

第 2 因子得点が表す値は、因子負荷量から負の値は距離の第 1 主成分得点が大きい馬を表していることがわかる。

2 因子が表す値から、距離以外の第 1 主成分得点が大きい馬で特に重賞の値が大きい馬が勝ちやすいことがわかる。図 4 より、第 2 因子得点が正の値をとる馬が多く有意な予想に用いるには適さないと判断する。

5.2.3　過去有馬記念からの結果の考察 1

以上より、主成分分析も因子分析も勝つ確率が高い馬が多過ぎて有意な予想に適さないと判断する。主成分分析の第 1 主成分得点と、因子分析の第 1 因子得点を比べると、第 1 主成分得点が正の値は全体的に高い値を表すが、第 1 因子得点が正の値は重賞の成績の第 1 主成分得点と右回りの第 1 主成分得点で表せる値になっている。

次に主成分分析の第 2 主成分得点と、因子分析の第 2 因子得点を比べると、どちらも負の値は距離の第 1 主成分得点が大きい馬を表している。

以上のことから、主成分分析と因子分析第２主成分得点と第２因子得点は同じ値を示しているが、第１主成分得点と第１因子得点は因子では距離の示す割合が減っている。これにより因子分析の特徴である、因子の軸により値が分かれて表せたことになる。

5.3 過去有馬記念を用いての結果２

ここでは、過去有馬記念の結果から有馬記念の勝因を探る。用いた３変量の構成は、

1. 距離の成績の 1 着に 10、２着に 5、３着に 2.5 を掛け、着外の回数を足したもの
2. 右回りの成績に１着に 5、２着に 2.5、３着に 1 を掛け、その中の中山の成績を２倍したもの
3. 重賞の成績に１着に 5、２着に 2.5、３着に 1 を掛け、その中のＧＩの成績を２倍にしたもの

である。§5.3.1 では、過去有馬記念の成績で主成分分析を行い、計算された固有ベクトルが勝因だと仮定し、2008 年有馬記念の出走馬に適用して、散布図を作成し、分析する。§5.3.2 では、中山競馬場の 2500 m の成績で因子分析を行い、計算された因子負荷量が勝因だと仮定し、因子得点係数を、2008 年有馬記念の出走馬に適用して、散布図を作成し、分析する。§5.3.3 に結果の考察を書く。使用した元のデータを表 7 に示す。過去有馬記念のデータは上から年代順に 1,2,3 着を示しており、それ以降は年代別の馬番の順に記載し、2008 年有馬記念は馬番順に記載している。この要素を標準化したものを用いた。

表 7: 過去有馬記念からの分析に使用した元データ 2

過去4年分の有馬記念の元データ					2008年有馬記念の元データ			
年代	馬番	3着率	距離の成績	中山競馬場と右回りの成績	馬番	3着率	距離の成績	中山競馬場と右回りの成績
2007年	3番	2.5	57.0	11.0	1番	0.0	17.5	28.5
2006年	4番	5.0	60.0	80.0	2番	0.0	7.5	2.5
2005年	10番	1.0	22.0	24.5	3番	7.0	37.5	27.5
2004年	1番	7.5	38.0	44.5	4番	12.0	46.0	5.0
2007年	7番	0.0	30.0	40.0	5番	0.0	22.5	10.0
2006年	1番	35.0	26.0	5.0	6番	0.0	45.5	23.0
2005年	6番	0.0	40.0	40.0	7番	7.5	15.5	10.5
2004年	9番	30.5	54.5	53.5	8番	10.0	27.0	18.5
2007年	4番	2.5	56.5	79.0	9番	2.0	53.5	74.5
2006年	5番	0.0	49.5	56.0	10番	22.5	89.5	33.5
2005年	14番	5.0	39.5	23.0	11番	0.0	37.0	27.0
2004年	6番	0.0	48.0	17.0	12番	0.0	14.5	25.5
2007年	1番	1.0	48.5	64.5	13番	5.0	40.0	55.0
2007年	2番	1.0	30.5	30.5	14番	20.5	69.0	17.0
2007年	5番	0.0	14.5	0.0				
2007年	6番	50.0	37.0	27.0				
2007年	8番	0.0	21.0	12.0				
2007年	9番	0.0	20.0	10.0				
2007年	10番	1.0	35.5	32.5				
2007年	11番	5.0	37.5	27.5				
2007年	12番	10.0	28.5	24.5				
2007年	13番	15.0	36.0	18.0				
2007年	14番	1.0	9.0	14.0				
2007年	15番	1.0	37.5	5.0				
2007年	16番	0.0	26.0	32.0				
2006年	2番	14.0	36.0	18.0				
2006年	3番	0.0	28.0	28.0				
2006年	6番	0.0	43.5	61.0				
2006年	7番	4.0	37.5	27.5				
2006年	8番	0.0	36.0	32.5				
2006年	9番	37.0	39.5	11.0				
2006年	10番	0.0	22.0	17.0				
2006年	11番	22.0	30.0	12.5				
2006年	12番	0.0	29.0	7.0				
2006年	13番	21.0	39.5	8.5				
2006年	14番	0.0	11.0	5.0				
2005年	1番	1.0	32.5	25.0				
2005年	2番	0.0	34.5	22.0				
2005年	3番	17.5	49.0	66.0				
2005年	4番	2.0	30.0	25.0				
2005年	5番	23.0	28.5	20.0				
2005年	7番	0.0	50.5	22.5				
2005年	8番	0.0	58.0	13.5				
2005年	9番	35.5	59.5	63.5				
2005年	11番	22.5	23.0	8.0				
2005年	12番	11.0	38.0	18.0				
2005年	13番	13.5	42.0	19.5				
2005年	15番	13.0	35.0	17.0				
2005年	16番	0.0	34.5	31.0				
2004年	2番	0.0	27.0	0.0				
2004年	3番	0.0	1.0	8.0				
2004年	4番	0.0	30.0	25.0				
2004年	5番	0.0	17.0	12.0				
2004年	7番	2.0	58.0	1.0				
2004年	8番	3.0	69.0	22.5				
2004年	10番	11.0	25.0	12.0				
2004年	11番	1.0	32.0	32.5				
2004年	12番	0.0	12.5	7.5				
2004年	13番	1.0	28.0	41.0				
2004年	14番	10.0	38.5	15.0				
2004年	15番	0.0	27.0	37.0				

5.3.1 過去有馬記念の主成分分析の結果 2

(2.10) 式で計算した、固有値と固有ベクトルは、固有値の大きい順番に、

$$\lambda = 1.539014, 0.987817, 0.473169$$

であり、固有ベクトルは

$$a = \begin{pmatrix} 0.252583 \\ 0.702767 \\ 0.665072 \end{pmatrix}, \begin{pmatrix} -0.944216 \\ 0.028905 \\ 0.328054 \end{pmatrix}, \begin{pmatrix} 0.211322 \\ -0.710833 \\ 0.670865 \end{pmatrix}$$

となる。固有値は 2 つで約 84 %なので 70 %を超えたので、第 2 主成分までで議論してよいことが確認された。

上記の固有値と固有ベクトルを勝因だと仮定し、2008 年有馬記念の出走馬の同じ要素に上記の固有ベクトルを掛け、(4.1) 式より主成分得点を計算した。計算された主成分得点を表 5 に載せる。表のデータは、右から過去有馬記念の出走馬の第 1 主成分得点、第 2 主成分得点、2008 年有馬記念の出走馬の第 1 主成分得点、第 2 主成分得点である。また過去有馬記念の上から年代順に 1,2,3 着を示しており、それ以降は年代別の馬番の順に記載し、2008 年有馬記念は馬番順に記載している。

表 8: 過去有馬記念からの分析の過去有馬記念と 2008 年有馬記念の主成分得点 2

過去有馬記念出走馬				2008年有馬記念出走馬		
年代	馬番	第1主成分得点	第2主成分得点	馬番	第1主成分得点	第2主成分得点
2007年	3番	0.515142	0.185218	1番	−0.740932	0.811775
2006年	4番	3.152638	1.179836	2番	−1.986161	0.342335
2005年	10番	−0.848127	0.470531	3番	0.102302	−0.070483
2004年	1番	0.825556	0.308408	4番	−0.255733	−1.090601
2007年	7番	0.087184	0.839959	5番	−1.238995	0.493708
2006年	1番	−0.567903	−2.699844	6番	−0.039976	0.752136
2005年	6番	0.602779	0.861166	7番	−1.189977	−0.461436
2004年	9番	2.506869	−1.421997	8番	−0.451973	−0.624131
2007年	4番	2.881135	1.363927	9番	2.116526	1.41185
2006年	5番	1.655364	1.158906	10番	2.508734	−1.869742
2005年	14番	0.0908	0.147437	11番	−0.169859	0.811135
2004年	6番	0.206273	0.479092	12番	−0.943731	0.755177
2007年	1番	1.925125	1.220711	13番	1.092625	0.669988
2007年	2番	−0.198832	0.592654	14番	1.197148	−1.931712
2007年	5番	−2.118913	0.113106			
2007年	6番	1.108293	−3.548005			
2007年	8番	−1.361699	0.335086			
2007年	9番	−1.483605	0.298266			
2007年	10番	0.129312	0.637956			
2007年	11番	0.145961	0.221269			
2007年	12番	−0.31185	−0.267593			
2007年	13番	−0.042035	−0.782186			
2007年	14番	−1.887718	0.260791			
2007年	15番	−0.73483	0.165085			
2007年	16番	−0.400439	0.69268			
2006年	2番	−0.064384	−0.698641			
2006年	3番	−0.438012	0.627523			
2006年	6番	1.521873	1.23293			
2006年	7番	0.123612	0.304814			
2006年	8番	0.132743	0.722562			
2006年	9番	0.383886	−2.734208			
2006年	10番	−1.134274	0.423954			
2006年	11番	−0.388402	−1.47515			
2006年	12番	−1.125089	0.265303			
2006年	13番	−0.061628	−1.440857			
2006年	14番	−2.123506	0.192432			
2005年	1番	−0.289165	0.501473			
2005年	2番	−0.313914	0.537211			
2005年	3番	2.372421	−0.130702			
2005年	4番	−0.395715	0.412626			
2005年	5番	−0.179594	−1.431755			
2005年	7番	0.528624	0.579816			
2005年	8番	0.598762	0.439575			
2005年	9番	3.228142	−1.655624			
2005年	11番	−0.896423	−1.609841			
2005年	12番	−0.028311	−0.443764			
2005年	13番	0.286559	−0.61812			
2005年	15番	−0.173465	−0.634566			
2005年	16番	0.002644	0.693357			
2004年	2番	−1.474419	0.139615			
2004年	3番	−2.533581	0.223274			
2004年	4番	−0.440413	0.579716			
2004年	5番	−1.567937	0.326603			
2004年	7番	0.203796	0.055615			
2004年	8番	1.549521	0.368413			
2004年	10番	−0.909623	−0.57543			
2004年	11番	−0.051147	0.630534			
2004年	12番	−1.958234	0.238987			
2004年	13番	0.041587	0.769522			
2004年	14番	−0.1304	−0.411207			
2004年	15番	−0.173014	0.781549			

次に散布図を図 5 に載せる。図の x 軸は第 1 主成分得点、y 軸は第 2 主成分得点、点の◆は過去有馬記念の出走馬の主成分得点、■は過去の 1,2,3 着の主成分得点、▲は 2008 年の出走馬の主成分得点である。

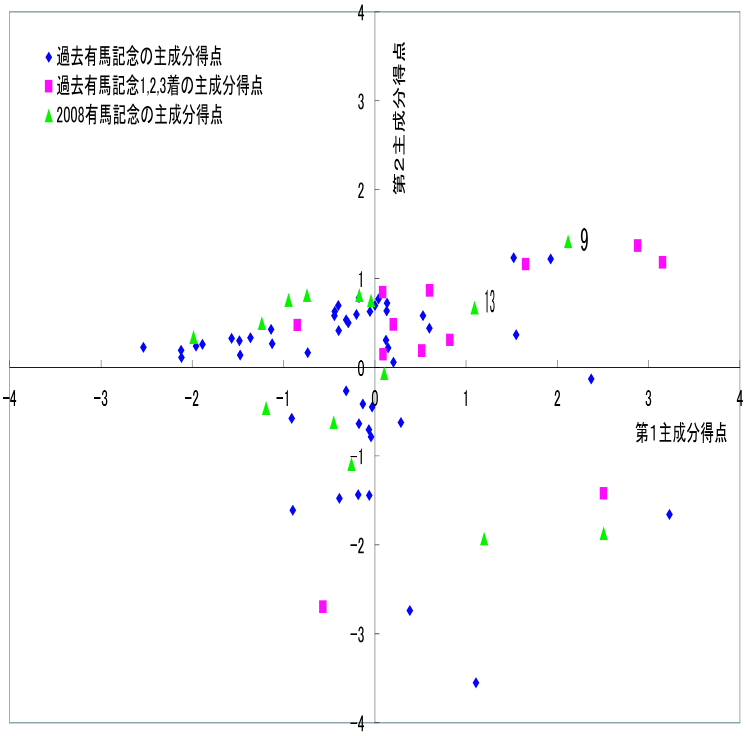

図 5: 過去有馬記念からの分析の過去有馬記念と 2008 年有馬記念の主成分得点の散布図 2

図 5 より第 1 主成分得点が正の値で第 2 主成分が正の値に相関が見られる。
第 1 主成分得点が表す値は、固有値から正の値は総合的に強い馬を表している。
第 2 主成分得点が表す値は、固有値から正の値は距離と重賞の成績が良い馬を表している。
２主成分から総合的に強く、距離と重賞の成績が良い馬が勝ちやすいことがわかる。図 5 より 2008 年有馬記念の出走馬の第 1 主成分得点が正の値で、第 2 主成分得点が正の値の馬は 9, 13 番である。

5.3.2 過去有馬記念を用いての因子分析の結果 2

(3.13) 式を因子負荷量の計算方法を用いて計算し、(3.23) 式によって回転された因子負荷量は

$$b = \begin{pmatrix} 0.039129 \\ 0.814888 \\ 0.871774 \end{pmatrix}, \begin{pmatrix} -0.991784 \\ -0.200343 \\ 0.066184 \end{pmatrix}$$

である。(3.30) 式で計算した因子得点係数は、

$$c = \begin{pmatrix} -0.063609 \\ 0.517921 \\ 0.612484 \end{pmatrix}, \begin{pmatrix} -0.977511 \\ -0.096501 \\ 0.133919 \end{pmatrix}$$

となる。上記の因子負荷量が勝因だと仮定し、因子得点係数を 2008 年有馬記念の要素に同じ因子得点係数を掛け、因子得点を計算する。表 6 に過去有馬記念の出走馬と 2008 年有馬記念の因子得点を載せる。表のデータは、右から過去有馬記念の出走馬の第 1 因子得点、第 2 因子得点、2008 年有馬記念の出走馬の第 1 因子得点、第 2 因子得点である。また過去有馬記念の上から年代順に 1,2,3 着を示しており、それ以降は年代別の馬番の順に記載し、2008 年有馬記念は馬番順に記載している。

表 9: 過去有馬記念からの分析の過去有馬記念と 2008 年有馬記念の因子得点 2

過去有馬記念出走馬				2008年有馬記念出走馬		
年代	馬番	第1因子得点	第2因子得点	馬番	第1因子得点	第2因子得点
2007年	3番	0.38877	0.146784	1番	−0.318846	0.92236
2006年	4番	2.72374	0.398008	2番	−1.406707	0.780072
2005年	10番	−0.495427	0.619932	3番	0.060358	−0.095218
2004年	1番	0.723798	0.08611	4番	−0.519111	−0.952689
2007年	7番	0.316262	0.759562	5番	−0.806981	0.767843
2006年	1番	−1.166436	−2.487198	6番	0.161745	0.759838
2005年	6番	0.696242	0.688763	7番	−1.020112	−0.186156
2004年	9番	1.512845	−1.956265	8番	−0.508041	−0.508867
2007年	4番	2.572426	0.631933	9番	2.020738	0.830035
2006年	5番	1.575494	0.734824	10番	1.351211	−2.323503
2005年	14番	0.098438	0.139444	11番	0.092203	0.825869
2004年	6番	0.25521	0.469226	12番	−0.487937	0.91406
2007年	1番	1.8072	0.715613	13番	1.037518	0.354357
2007年	2番	0.021908	0.602247	14番	0.343961	−2.088001
2007年	5番	−1.568387	0.586001			
2007年	6番	−0.120257	−3.706631			
2007年	8番	−0.932696	0.624972			
2007年	9番	−1.035478	0.617887			
2007年	10番	0.276683	0.581013			
2007年	11番	0.168206	0.185475			
2007年	12番	−0.299093	−0.204509			
2007年	13番	−0.252797	−0.736101			
2007年	14番	−1.329516	0.637605			
2007年	15番	−0.538101	0.372084			
2007年	16番	−0.094866	0.731222			
2006年	2番	−0.247169	−0.649609			
2006年	3番	−0.148438	0.688732			
2006年	6番	1.509466	0.812716			
2006年	7番	0.173834	0.271967			
2006年	8番	0.30131	0.663964			
2006年	9番	−0.470368	−2.713265			
2006年	10番	−0.732738	0.653304			
2006年	11番	−0.698338	−1.338013			
2006年	12番	−0.790672	0.53292			
2006年	13番	−0.461297	−1.347112			
2006年	14番	−1.53942	0.646194			
2005年	1番	−0.080251	0.549134			
2005年	2番	−0.095803	0.600218			
2005年	3番	1.781922	−0.704408			
2005年	4番	−0.180875	0.480342			
2005年	5番	−0.518023	−1.360766			
2005年	7番	0.528361	0.49048			
2005年	8番	0.521818	0.373638			
2005年	9番	1.998614	−2.353296			
2005年	11番	−1.112902	−1.363571			
2005年	12番	−0.154288	−0.404296			
2005年	13番	0.032221	−0.63822			
2005年	15番	−0.31193	−0.563121			
2005年	16番	0.195725	0.66396			
2004年	2番	−1.093412	0.497502			
2004年	3番	−1.822224	0.73824			
2004年	4番	−0.169618	0.653325			
2004年	5番	−1.084688	0.653291			
2004年	7番	0.105662	0.112124			
2004年	8番	1.21444	0.100028			
2004年	10番	−0.842614	−0.354752			
2004年	11番	0.143689	0.605792			
2004年	12番	−1.401443	0.65328			
2004年	13番	0.267029	0.694313			
2004年	14番	−0.226837	−0.342592			
2004年	15番	0.105092	0.759554			

次に散布図を図 6 に載せる。図の x 軸は第 1 因子得点、y 軸は第 2 因子得点、点の◆は過去有馬記念の出走馬の因子得点、■は過去の 1,2,3 着の因子得点、▲は 2008 年の出走馬の主成分得点である。

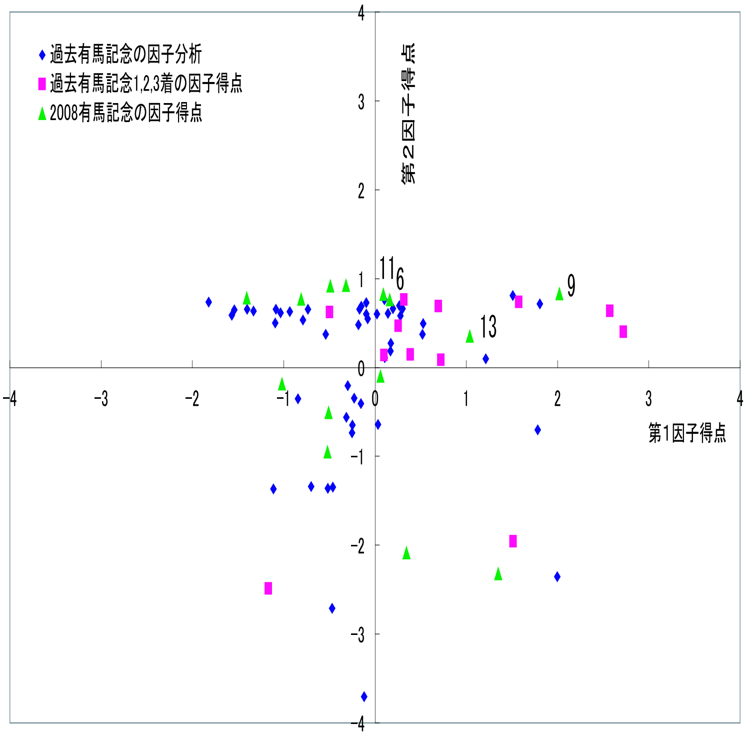

図 6: 過去有馬記念からの分析の過去有馬記念と 2008 年有馬記念の因子得点の散布図 2

図 6 より第 1 因子得点が正の値で第 2 因子得点が正の値に相関が見られる。

第 1 因子得点が表す値は、因子負荷量から正の値は重賞の成績がよく右回りの中山競馬場が得意な馬を表していることがわかる。

第 2 因子得点が表す値は、因子負荷量から正の値は重賞の成績が良い馬を表していることがわかる。

2 因子から、因子負荷量の重賞の成績を特に大きい値をとる馬がよい成績をのこすことがわかる。図 6 より、2008 年有馬記念の出走馬の第 1 因子得点が正の値で、第 2 因子得点が正の値をとる馬は $6, 9, 11, 13$ 番である。

5.3.3　過去有馬記念からの結果の考察 2

以上より、主成分分析より、$9, 13$ 番が勝ち易いといえ、因子分析より $6, 9, 11, 13$ 番が勝ち易いといえる。特に、どちらの分析にも出ている $9, 13$ 番は勝つ確率が高いといえる。

主成分分析の第 1 主成分得点と、因子分析の第 1 因子得点を比べると、第 1 主成分得点が正の値は総合力を表すが、第 1 因子得点が正の値は重賞の成績がよく右回りの中山競馬場が得意な馬で表せるようになっている。ただ因子分析の因子負荷量の結果の結果も距離の成績を正の値をとっているため、お互いの得点を示す値は大きく変わらないといえる。

次に主成分分析の第２主成分得点と、因子分析の第２因子得点を比べると、お互い同じ重賞の成績が良い馬が正の値を表している。

以上のことから、主成分分析と因子分析第２主成分得点と第２因子得点は同じ値を示しているが、第１主成分得点と第１因子得点は因子では距離の示す割合が減っている。これにより因子分析の特徴である、因子の軸により値が分かれて表せたことになる。

5.4 研究結果と実際の結果

以上の研究結果より、2008 年有馬記念の出走馬 14 頭中二つの結果からの分析で合計６頭の馬に相関が見れた。ここで少し絞り込むために、どちらの分析にも出ている 6, 9, 13 番が有馬記念で勝つ確率の高い馬であると予想する。予想結果のレースでの人気は次の通りである、6 番 10 番人気、9 番は 4 番人気、13 番は 1 番人気であった。人気を集める馬が２頭と穴馬が１頭上がったことになる。

次に 12 月 28 日に行なわれたの有馬記念のレース結果を下に示す。

表 10: 2008 年有馬記念のレース結果

着順		馬名	払い戻し金		
1着	13番	ダイワスカーレット			
2着	14番	アドマイヤモナーク			
3着	6番	エアシェイディ	単勝	13番	260円
4着	11番	ドリームジャーニー	複勝	6番	600円
5着	8番	スクリーンヒーロー		13番	130円
6着	7番	アルナスライン		14番	2,280円
7着	1番	カワカミプリンセス	枠連	8-8	18,640円
8着	9番	メイショウサムソン	馬連	13-14	29,490円
9着	5番	フローテーション	ワイド	6-13	1,360円
10着	2番	ベンチャーナイン		6-14	28,200円
11着	3番	コスモバルク		13-14	7,160円
12着	10番	マツリダゴッホ	馬単	13-14	33,490円
13着	4番	エアジパング	3連複	6-13-14	192,500円
14着	12番	アサクサキングス	3連単	13-14-6	985,580円

であった。研究の結果である 6, 13 番は馬券圏内である３着以内に入ったが、9 番がこなかったことになる。

6　まとめと展望

競馬の勝因は馬のコース、距離の得意、不得意、クラス分けされた力の違いである程度の予想は可能である事が分析により明確になった。これにより、競馬新聞に載っている、データで予想が可能であるといえる。

主成分分析と因子分析ともに、極端に違う形をした散布図はできなかった。これは、要素を 3 要素と絞って分析ため、このような結果になったのではないかと考える。しかし主成分分析の固有ベクトル、因子分析の因子負荷量の結果を見比べると、同じような値を示しているが、主成分分析の主成分得点では、全ての要素の持つ割合が高いのに対し、因子分析の因子得点では、１つもしくは２つの要素が高い割合の値に変わっていた。この結果は主成分分析と因子分析の特徴が出たものになった。

次に本研究の結果の馬について、レースに勝った 13 番は本研究の結果でも高い勝因があることが確認された。同じく 6 番も、馬券圏内である 3 着に入選した。しかし、同じように高い勝因があった 9 番は、8 着と大敗している。また、２着に入選した 14 番は、中山競馬場の成績から分析した結果からは、勝因が高いとでたが、過去有馬記念からの分析からは、分析できなかった。9 番の敗因と 14 番の入選したことを、解析することが、今後の課題としてあげられる。

今後の展望として、本研究では主成分分析、因子分析を 3 要素で研究を進めてきたが、この要素を増やして研究すると 14 番に相関を持たせることができるかもしれないと考える。それには、過去有馬記念の結果からも、毎年１頭は予想不可能な馬もでているので、その馬について相関を持たせることが 14 番に相関を持たせる近道だと考える。

本研究で得られた結果より、用いた要素の中で一番勝因として大きいと出たものは、「右回りの成績」と「中山競馬場が得意」である。これにより、馬には「右回りが得意」、「左回りが得意」、「右回りでも特に中山競馬場が得意」という風にコースによって得意、不得意があることがわかり、それが最も大きく作用していることがわかった。次に重賞の成績である。これはいくらコースが得意でも馬に力がないと勝てないことを示していると考える。最後に距離の要素である。馬にとって、上記の二つからすると、距離は問題ではないということがいえる。

最後に、研究の結果の馬券は外れたが、今回用いたデータでもある程度予想可能だとわかり、良い結果であった。

参考文献

[1] 三土修平著　初歩からの多変量分析　日本評論社 (1997)

主成分分析のソースファイル

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>

double hensa(double h)
{
return(sqrt(h));
}

double kb(double K,double B)
{
return(K*B);
}
struct data{
double x;
double y;
double z;
};
typedef struct data DATA;

struct point{
double x;
double y;
double z;
};
typedef struct point POINT;

struct kyou{
double x;
double y;
double z;
};
typedef struct kyou KYOU;

int main(void)
{
DATA d[300];
POINT p[300];
KYOU b[300];
FILE *fp;
```

```
double x,y,z;
int i=0,j,k;
double sumX=0,sumY=0,sumZ=0;
double aveX=0,aveY=0,aveZ=0;
double sumdX=0,sumdY=0,sumdZ=0;
double avedX=0,avedY=0,avedZ=0;
double sumkX=0,sumkY=0,sumkZ=0;
double avekX=0,avekY=0,avekZ=0;
double kyoX,kyoY,kyoZ;
double bun=1.0;
double A1,A2,A3,Q,R,si;
double l1,l2,l3;
double lmax,lmin;
double a11,a21,a31;
double a12,a22,a32;
double a13,a23,a33;
double c1,c2,c3;
double D,DD,DDD;
double syu1,syu2,syu3;
int m;
fp=fopen("[入力ファイル名].txt","r");
if(fp == NULL) exit(1);
while(fscanf(fp,"%lf %lf %lf",&x,&y,&z)!=EOF){
d[i].x=x;
d[i].y=y;
d[i].z=z;
sumX+=x;
sumY+=y;
sumZ+=z;

i++;
}
fclose(fp);
aveX=sumX/i;
aveY=sumY/i;
aveZ=sumZ/i;

for(j=0;j<i;j++){
d[j].x-=aveX;
d[j].y-=aveY;
d[j].z-=aveZ;
```

```
p[j].x=kb(d[j].x,d[j].x);
p[j].y=kb(d[j].y,d[j].y);
p[j].z=kb(d[j].z,d[j].z);
b[j].x=kb(d[j].x,d[j].y);
b[j].y=kb(d[j].y,d[j].z);
b[j].z=kb(d[j].z,d[j].x);
sumdX+=p[j].x;
sumdY+=p[j].y;
sumdZ+=p[j].z;
sumkX+=b[j].x;
sumkY+=b[j].y;
sumkZ+=b[j].z;
}
avedX=sumdX/j;
avedY=sumdY/j;
avedZ=sumdZ/j;
avekX=sumkX/j;
avekY=sumkY/j;
avekZ=sumkZ/j;
kyoX=avekX/(hensa(avedX)*hensa(avedY));
kyoY=avekY/(hensa(avedY)*hensa(avedZ));
kyoZ=avekZ/(hensa(avedZ)*hensa(avedX));
   for(k=0;k<j;k++){
d[k].x/=hensa(avedX);
d[k].y/=hensa(avedY);
d[k].z/=hensa(avedZ);
}
A1=-3.0*bun;
A2=3*bun-kyoX*kyoX-kyoY*kyoY-kyoZ*kyoZ;
A3=bun*(kyoX*kyoX+kyoY*kyoY+kyoZ*kyoZ)-bun*bun*bun-2*kyoX*kyoY*kyoZ;
Q=(A1*A1-3*A2)/9;
R=(2*A1*A1*A1-9.0*A1*A2+27*A3)/54;
si=acos(R/sqrt(Q*Q*Q));
l1=-2*hensa(Q)*cos(si/3)-A1/3;
l2=-2*hensa(Q)*cos((si+2*M_PI)/3)-A1/3;
l3=-2*hensa(Q)*cos((si+4*M_PI)/3)-A1/3;
if(l1<l2){
lmax=l2;
l2=l1;
l1=lmax;
} if(l2<l3){
```

```
lmin=l2;
l2=l3;
l3=lmin;
} if(l1<l2){
lmax=l2;
l2=l1;
l1=lmax;}
c1=(kyoX*(l1-bun)+kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l1-bun));
c2=-((kyoX*kyoX*(l1-bun)+kyoX*kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l1-bun))+kyoY)/(bun-l1);
c3=1.0;
D=sqrt(c1*c1+c2*c2+c3*c3);
a11=c1/D;
a21=c2/D;
a31=c3/D;
c1=(kyoX*(l2-bun)+kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l2-bun));
c2=-((kyoX*kyoX*(l2-bun)+kyoX*kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l2-bun))+kyoY)/(bun-l2);
c3=1.0;
DD=sqrt(c1*c1+c2*c2+c3*c3);
a12=c1/DD;
a22=c2/DD;
a32=c3/DD;
c1=(kyoX*(l3-bun)+kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l3-bun));
c2=-((kyoX*kyoX*(l3-bun)+kyoX*kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l3-bun))+kyoY)/(bun-l3);
c3=1.0;
DDD=sqrt(c1*c1+c2*c2+c3*c3);
a13=c1/DDD;
a23=c2/DDD;
a33=c3/DDD;
fp=fopen("[出力ファイル名].txt","w");
fprintf(fp,"固有値\nl1:%lf\tl2:%lf\tl3:%lf\n",l1,l2,l3);
fprintf(fp,"寄与率 \nl1:%.1f％\tl2:%.1f％\tl3:%.1f％\n",l1/(l1+l2+l3)*100,l2/(l1+l2
fprintf(fp,"累積寄与率\n ν:%.1f％\n",(l1+l2)/(l1+l2+l3)*100);
fprintf(fp,"固有ベクトル\na11:%lf\ta12:%lf\ta13:%lf\n",a11,a12,a13);
fprintf(fp,"a21:%lf\ta22:%lf\ta23:%lf\n",a21,a22,a23);
fprintf(fp,"a31:%lf\ta32:%lf\ta33:%lf\n",a31,a32,a33);
fclose(fp);
fp=fopen("[出力ファイル名].xls","w");
fprintf(fp,"第1主成分\t第2主成分\t第3主成分\n");
for(m=0;m<k;m++){
syu1=d[m].x*a11+d[m].y*a21+d[m].z*a31;
syu2=d[m].x*a12+d[m].y*a22+d[m].z*a32;
```

```
syu3=d[m].x*a13+d[m].y*a23+d[m].z*a33;
fprintf(fp,"%lf\t%lf\t%lf\n",syu1,syu2,syu3);
}
fclose(fp);
return(0);
}
```

因子分析のソースファイル

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#define PI  3.141592653589793

double hensa(double h){
return(sqrt(h));
}
double kb(double K,double B){
return(K*B);
}
struct data{
double x;
double y;
double z;
};
typedef struct data DATA;

struct point{
double x;
double y;
double z;
};
typedef struct point POINT;

struct kyou{
double x;
double y;
double z;
};
typedef struct kyou KYOU;

int main(void)
{
DATA d[300];
POINT p[300];
KYOU b[300];
FILE *fp;
double x,y,z,bun=1.0;
int i=0,j,k,l;
```

```
double sumX=0,sumY=0,sumZ=0;
double aveX=0,aveY=0,aveZ=0;
double sumdX=0,sumdY=0,sumdZ=0;
double avedX=0,avedY=0,avedZ=0;
double sumkX=0,sumkY=0,sumkZ=0;
double avekX=0,avekY=0,avekZ=0;
double kyoX,kyoY,kyoZ;
double A1,A2,A3,Q,R,si;
double l1,l2,l3;
double ll1,ll2,ll3;
double lmax,lmin;
int m,mmax;
double bb11,bb21,bb31;
double bb12,bb22,bb32;
double d1,d2,d3;
double D,DD;
double sita,s1,s2,ss1,ss2,S;
double Smax=0,Ssita;
double B1,B2,B3;
double BB1,BB2,BB3;
double aa1,aa2,aa3;
double bb1,bb2,bb3;
double b11,b21,b31;
double b12,b22,b32;
double inbun1,inbun2,inbun3;
double ind1,ind2,ind3;
double nind1,nind2,nind3;
double sa;
double gkyoX,gkyoY,gkyoZ,detA;
double gbun1,gbun2,gbun3;
double c11,c21,c31;
double c12,c22,c32;
double insi1,insi2;
fp=fopen("[入力ファイル名].txt","r");
if(fp == NULL) exit(1);
while(fscanf(fp,"%lf %lf %lf",&x,&y,&z)!=EOF){
d[i].x=x;
d[i].y=y;
d[i].z=z;
sumX+=x;
sumY+=y;
```

```
sumZ+=z;
i++;
}
fclose(fp);
aveX=sumX/i;
aveY=sumY/i;
aveZ=sumZ/i;
for(j=0;j<i;j++){
d[j].x-=aveX;
d[j].y-=aveY;
d[j].z-=aveZ;
p[j].x=kb(d[j].x,d[j].x);
p[j].y=kb(d[j].y,d[j].y);
p[j].z=kb(d[j].z,d[j].z);
b[j].x=kb(d[j].x,d[j].y);
b[j].y=kb(d[j].y,d[j].z);
b[j].z=kb(d[j].z,d[j].x);
sumdX+=p[j].x;
sumdY+=p[j].y;
sumdZ+=p[j].z;
sumkX+=b[j].x;
sumkY+=b[j].y;
sumkZ+=b[j].z;
}
avedX=sumdX/j;
avedY=sumdY/j;
avedZ=sumdZ/j;
avekX=sumkX/j;
avekY=sumkY/j;
avekZ=sumkZ/j;
kyoX=avekX/(hensa(avedX)*hensa(avedY));
kyoY=avekY/(hensa(avedY)*hensa(avedZ));
kyoZ=avekZ/(hensa(avedZ)*hensa(avedX));
for(k=0;k<j;k++){
d[k].x/=hensa(avedX);
d[k].y/=hensa(avedY);
d[k].z/=hensa(avedZ);
}
ind1=0;
ind2=0;
ind3=0;
```

```
for(l=1;;l++){
inbun1=1-ind1*ind1;
inbun2=1-ind2*ind2;
inbun3=1-ind3*ind3;

A1=-(inbun1+inbun2+inbun3);
A2=inbun1*inbun2+inbun2*inbun3+inbun3*inbun1-kyoX*kyoX-kyoY*kyoY-kyoZ*kyoZ;
A3=inbun1*kyoY*kyoY+inbun2*kyoZ*kyoZ+inbun3*kyoX*kyoX-inbun1*inbun2*inbun3-2*kyoX*k
Q=(A1*A1-3*A2)/9;
R=(2*A1*A1*A1-9.0*A1*A2+27*A3)/54;
si=acos(R/sqrt(Q*Q*Q));
l1=-2*hensa(Q)*cos(si/3)-A1/3;
l2=-2*hensa(Q)*cos((si+2*M_PI)/3)-A1/3;
l3=-2*hensa(Q)*cos((si+4*M_PI)/3)-A1/3;
if(l1<l2){
lmax=l2;
l2=l1;
l1=lmax;
} if(l2<l3){
lmin=l2;
l2=l3;
l3=lmin;
} if(l1<l2){
lmax=l2;
l2=l1;
l1=lmax;}
d1=(kyoX*(l1-inbun3)+kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l1-inbun1));
d2=-((kyoX*kyoX*(l1-inbun3)+kyoX*kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l1-inbun1))
+kyoY)/(inbun2-l1);
d3=1.0;
D=sqrt(d1*d1+d2*d2+d3*d3);
b11=d1/D;
b21=d2/D;
b31=d3/D;

d1=(kyoX*(l2-inbun3)+kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l2-inbun1));
d2=-((kyoX*kyoX*(l2-inbun3)+kyoX*kyoY*kyoZ)/(kyoX*kyoZ+kyoY*(l2-inbun1))
+kyoY)/(inbun2-l2);
d3=1.0;
DD=sqrt(d1*d1+d2*d2+d3*d3);
```

```
b12=d1/DD;
b22=d2/DD;
b32=d3/DD;
aa1=sqrt(l1)*b11;
aa2=sqrt(l1)*b21;
aa3=sqrt(l1)*b31;
bb1=sqrt(l2)*b12;
bb2=sqrt(l2)*b22;
bb3=sqrt(l2)*b32;

nind1=1-(aa1*aa1+bb1*bb1);
nind2=1-(aa2*aa2+bb2*bb2);
nind3=1-(aa3*aa3+bb3*bb3);
sa=(ind1-nind1)*(ind1-nind1)+(ind2-nind2)*(ind2-nind2)+(ind3-nind3)*(ind3-nind3);
if(sa<0.00001){
l1=ll1;
l2=ll2;
l3=ll3;
break;
}
ll1=l1;
ll2=l2;
ll3=l3;
ind1=nind1;
ind2=nind2;
ind3=nind3;
bb11=aa1;
bb21=aa2;
bb31=aa3;
bb12=bb1;
bb22=bb2;
bb32=bb3;
}
for(m=0;m<=90;m++){
sita=(PI/180)*m;
B1=(bb11*cos(sita)+bb12*sin(sita))/sqrt(inbun1);
B2=(bb21*cos(sita)+bb22*sin(sita))/sqrt(inbun2);
B3=(bb31*cos(sita)+bb32*sin(sita))/sqrt(inbun3);
BB1=(-bb11*sin(sita)+bb12*cos(sita))/sqrt(inbun1);
BB2=(-bb21*sin(sita)+bb22*cos(sita))/sqrt(inbun2);
BB3=(-bb31*sin(sita)+bb32*cos(sita))/sqrt(inbun3);
```

```
s1=B1*B1*B1*B1+B2*B2*B2*B2+B3*B3*B3*B3;
s2=BB1*BB1*BB1*BB1+BB2*BB2*BB2*BB2+BB3*BB3*BB3*BB3;
ss1=B1*B1+B2*B2+B3*B3;
ss2=BB1*BB1+BB2*BB2+BB3*BB3;

S=(3*s1-ss1*ss1)+
(3*s2-ss2*ss2);
if(S>Smax){
Smax=S;
Ssita=sita;
}
}
b11=bb11*cos(Ssita)+bb12*sin(Ssita);
b21=bb21*cos(Ssita)+bb22*sin(Ssita);
b31=bb31*cos(Ssita)+bb32*sin(Ssita);
b12=-bb11*sin(Ssita)+bb12*cos(Ssita);
b22=-bb21*sin(Ssita)+bb22*cos(Ssita);
b32=-bb31*sin(Ssita)+bb32*cos(Ssita);
gbun1=bun*bun-kyoY*kyoY;
gbun2=bun*bun-kyoZ*kyoZ;
gbun3=bun*bun-kyoX*kyoX;
gkyoX=kyoY*kyoZ-bun*kyoX;
gkyoY=kyoZ*kyoX-bun*kyoY;
gkyoZ=kyoX*kyoY-bun*kyoZ;
detA=bun*gbun1+kyoX*gkyoX+kyoZ*gkyoZ;
gbun1=gbun1/detA;
gbun2=gbun2/detA;
gbun3=gbun3/detA;
gkyoX=gkyoX/detA;
gkyoY=gkyoY/detA;
gkyoZ=gkyoZ/detA;
c11=gbun1*b11+gkyoX*b21+gkyoZ*b31;
c21=gkyoX*b11+gbun2*b21+gkyoY*b31;
c31=gkyoZ*b11+gkyoY*b21+gbun3*b31;
c12=gbun1*b12+gkyoX*b22+gkyoZ*b32;
c22=gkyoX*b12+gbun2*b22+gkyoY*b32;
c32=gkyoZ*b12+gkyoY*b22+gbun3*b32;
fp=fopen("[出力ファイル名].txt","w");
fprintf(fp,"回転後の因子負荷量\nb11:%lf\tb12:%lf\n",b11,b12);
fprintf(fp,"b21:%lf\tb22:%lf\n",b21,b22);
fprintf(fp,"b31:%lf\tb32:%lf\n",b31,b32);
```

```
fprintf(fp,"回転後の因子得点の係数\nc11:%lf\tc12:%lf\n",c11,c12);
fprintf(fp,"c21:%lf\tc22:%lf\n",c21,c22);
fprintf(fp,"c31:%lf\tc32:%lf\n",c31,c32);
fclose(fp);
fp=fopen("[出力ファイル名].xls","w");
fprintf(fp,"回転後\n");
for(m=0;m<k;m++){
insi1=d[m].x*c11+d[m].y*c21+d[m].z*c31;
insi2=d[m].x*c12+d[m].y*c22+d[m].z*c32;
fprintf(fp,"%lf\t%lf\n",insi1,insi2);
}
fclose(fp);
return(0);
}
```