浜松医科大学紀要　一般教育　第20号（2006）

競馬データにみられる統計的偏りについて（3）

野田　明男
（総合人間科学講座・数学）

On the Statistical Bias Found in the Horse Racing Data (3)

Akio NODA
Integrated Human Sciences・Mathematics

Abstract: This is a continuation of the author's previous papers [2] and [3] . Our approach based on exchangeable random variables t_i (i =1, 2, 3, 4) was introduced in [3] and is improved in § 1. By separating all horse racings into the two categories of dirt and turf (such a separation was suggested by a referee of [3]) , we are able to get better results on the statistical bias found in the racing files [4] . Here we take up only racings of 16 participants carried out on these racetracks: Chukyo, Hanshin, Kyoto, Nakayama and Tokyo.

The present data analysis thus performed leads us to investigate some characteristics for each racetrack mentioned above. In fact, we first make up a lot of contingency tables naturally arising from our approach, and by appealing to the familiar chi-square test ([1]) , we see how large the deviation is between the empirical and expected frequencies. We are then in a position to report what kinds of differences are observed among these racetracks and also between the two categories of dirt and turf on the same racetrack. The details are in § 2.

Key words: chi-square test, contingency table, exchangeable random variables.

§1. 序

著者の論文[2][3]に引き続いて、中央競馬レース成績[4]で見出される統計的偏りを考察する。5つの競馬場(イ)中京(ロ)阪神(ハ)京都(ニ)中山(ホ)東京におけるレース成績を比較検討し、各競馬場の「個性」を解明したい、これが著者の研究目標である。この論文では[3]のアプローチの不十分な点を改善した上で、「芝」と「ダート」の2つに分けて、$m=16$ 頭のレース結果を分析する。[3]のレフェリーから示唆されたこの分離分割のおかげで、偏り方に関する競馬場間の差異(および類似性)が一層鮮明になる。互いに強め合う方向に、大きな偏りをはっきりと示す阪神の芝とダート、いろいろな面で偏りを示すダートの一方で、全く偏りを見せぬ芝をもつ京都、そして互いに打ち消し合う方向に、少数の分類項目でしか偏りを見せぬ中山の芝とダート、これら3者が特にあざやかな対比を形作る。次節において、各競馬場毎にダートと芝のレース成績を分析した結果をそれぞれ詳述する。加えて、ダートと芝の偏り方の異同について、χ^2 統計量の値の変化に着目して調べた結果を報告する。

[4]に記載されている n 回のレース結果は、出走馬の馬番を表す有限母集団 $\{1, 2, \cdots, m\}$ から、上位3頭の無作為抽出が n 回繰り返されたものとみなす。これが偏りを測る基準となる帰無仮説 H_0 である。さて、1着 x_1、2着 x_2、3着 x_3 と記すとき、(x_1, x_2, x_3) の確率分布は H_0 の下で交換可能である([5]参照)。しかしながら、3連単の立場でレース結果を調べて行くと、分類項目が多くなって、統計的偏りは検出しにくい傾向がある。くじ引きで馬番が決まるので、偶然の要素がかなり大きく寄与するものと思われる。われわれは3連複の立場に移って、レース成績を整理し直す。そのため、x_1, x_2, x_3 の順序統計量 $x_{(1)} < x_{(2)} < x_{(3)}$ 間の差

$$t_1 = x_{(1)},\ t_2 = x_{(2)} - x_{(1)},\ t_3 = x_{(3)} - x_{(2)},\ t_4 = m+1-x_{(3)}$$

を基本統計量として採用する。$t_1+t_2+t_3+t_4 = m+1$ で和が一定の (t_1, t_2, t_3, t_4) の確率分布は、順序統計量に移って一端失った交換可能性を回復する。[3]と同じアプローチにより、基本統計量 t_i 達の大小関係および t_i のとる値に基づいて、レース結果を分類し、対応する分割表(度数分布表)を作成して、χ^2 検定([1]参照)を実行する。その結果を比較検討し、まとめたものが§2に他ならぬ。

[3]におけるデータ分析の不十分さを改善するため、今回新たに導入したやり方を述べて、次節への準備とする。まず t_1, t_2, t_3, t_4 の中から2つ選んで t_i, t_j $(i < j)$ とし、

$$X_0 = \{t_i = t_j\} \quad X_1 = \{t_i < t_j\} \quad X_2 = \{t_i > t_j\}$$

と定義する。即ち、[3]の内枠・外枠問題を意識して定めた $(i, j) = (2, 3)$ の $X = A$, $(1, 3)$ の B, $(2, 4)$ の C, $(1, 4)$ の D の4種類に加えて、$(i, j) = (1, 2)$ の $X = E$, $(3, 4)$ の F の2種類を新たに考える。具体的に記すと次の通り。

$$A_0 = \{t_2 = t_3\} \quad A_1 = \{t_2 < t_3\} \quad A_2 = \{t_2 > t_3\}$$

$$B_0 = \{t_1 = t_3\} \quad B_1 = \{t_1 < t_3\} \quad B_2 = \{t_1 > t_3\}$$

$$C_0 = \{t_2 = t_4\} \quad C_1 = \{t_2 < t_4\} \quad C_2 = \{t_2 > t_4\}$$

$D_0=\{t_1=t_4\}\quad D_1=\{t_1<t_4\}\quad D_2=\{t_1>t_4\}$

$E_0=\{t_1=t_2\}\quad E_1=\{t_1<t_2\}\quad E_2=\{t_1>t_2\}$

$F_0=\{t_3=t_4\}\quad F_1=\{t_3<t_4\}\quad F_2=\{t_3>t_4\}$

さらに、$A\sim F$ 6種類の分割の中から2つ選んでX, Yとし、積事象 $X_i\cap Y_j\,(i, j=0, 1, 2)$ を構成する。こうしてできる(6＋15)通りの分割がこの論文において第1の役割を担う分類項目である。このときH_0の下での確率分布は、t_i達の交換可能性を考慮して、XとYに用いられるt_i達に重複がなければ($A_i\cap D_j$, $B_i\cap C_j$, $E_i\cap F_j$ の3通り)[3]の命題3を、重複が1つある場合($A_i\cap B_j$など残り12通り)は[2]の命題4を参照すればよい。

次に、t_1, t_2, t_3, t_4 の順序統計量 $t_{(1)}\leqq t_{(2)}\leqq t_{(3)}\leqq t_{(4)}$ のうち、両端をなす最小値 $t_{(1)}$ と最大値 $t_{(4)}$ に着目して、次の4通りの分類(第2の役割を担う)を行う。以下 $t_{(k)}=t_{i_k}\,(k=1, 2, 3, 4)$ と書く。

(a) $t_{(1)}<t_{(2)}$ となる場合、最小値の位置 i_1 に加えて、その値 t_{i_1} との組 (i_1, t_{i_1})

(b) $t_{(3)}<t_{(4)}$ となる場合、最大値の位置 i_4 に加えて、その値 t_{i_4} との組 (i_4, t_{i_4})

(c) $t_{(1)}=t_{(2)}<t_{(3)}$ となる場合、その位置 (i_1, i_2) に加えて、その値との組 (i_1, i_2, t_{i_1})

(d) $t_{(1)}<t_{(2)}\leqq t_{(3)}<t_{(4)}$ となる場合、両端をなす位置の組 (i_1, i_4)

$m=16$ のとき、交換可能性のおかげで、次のように確率計算される。i_1, i_4 は1から4までの整数値 $(i_1\neq i_4)$, (i_1, i_2) は $1\leqq i_1<i_2\leqq 4$ を満たす。

(a) $P(i_1, t_{i_1}=1)=\dfrac{66}{400}, P(i_1, t_{i_1}=2)=\dfrac{28}{400}, P(i_1, t_{i_1}=3)=\dfrac{6}{400}$。

(b) $P(i_4, t_{i_4}\leqq 6)=\dfrac{16}{508}, P(i_4, t_{i_4}=7)=\dfrac{27}{508}, P(i_4, t_{i_4}=8)=\dfrac{28}{508}, P(i_4, t_{i_4}=9)=\dfrac{21}{508},$

$P(i_4, t_{i_4}=10)=\dfrac{15}{508}, P(i_4, t_{i_4}\geqq 11)=\dfrac{20}{508}$。

(c) $P(i_1, i_2, t_{i_1}=1)=\dfrac{1}{12}, P(i_1, i_2, t_{i_1}=2)=\dfrac{1}{18}, P(i_1, i_2, t_{i_1}=3)=\dfrac{1}{36}$。

(d) $P(i_1, i_4)=\dfrac{1}{12}$ の一様分布。

最後に[3]で採用した上記以外の分類項目を取りあげる。対応する確率分布表はすべて交換可能性から容易に従う([3]参照)。

§2. レース成績から競馬場の個性を引き出す試み

§1で述べたアプローチで、$m = 16$ 頭のレースをダートと芝に分けて、データ分析（χ^2 検定）を行った結果を5つの競馬場毎に詳述する。有意水準は5%とし、帰無仮説 H_0 が棄却されるケースを主として取りあげて行く。さらに、ダートと芝で偏り方にどんな違いがあるか、調べることにする。

（i） 両者を合わせると χ^2 値が大きくなる（互いに強め合う向きの偏りを示す）ケース；

（ii） 両者を合わせると χ^2 値が小さくなる（互いに打ち消し合うような偏りを示す）ケース；

（iii） 上記どちらでもない（両者の χ^2 値の中間におさまる）ケース。

このような χ^2 値の変化に連動する P 値の変化を調べ、興味深い点を報告する。

（イ）中京競馬場：[4]に記載されているレース数は、ダートの $n_1 = 405$ と芝の $n_2 = 268$ であり、全レース $(N = 1524)$ 中それぞれ26.6%と17.6%を占める。

ダート（1）まずB, D, E において、χ^2 値が大きく、P 値はそれぞれ0.1%未満、1～2.5%、0.1%未満になる。いずれも高頻度で、t_1 の値が他の t_i に比べて大きくなる。B, D, E を含む積事象を調べると、$A_i \cap D_j$ 以外すべて5%未満になる。特に、0.1%未満の P 値を示すのは $A_i \cap B_j$, $A_i \cap E_j$, $B_i \cap C_j$, $B_i \cap E_j$, $B_i \cap F_j$ の5つであり、0.1～0.5%になるのは $B_i \cap D_j$ と $E_i \cap F_j$ 、1～2.5%になるのは $D_i \cap E_j$ と $D_i \cap F_j$ である。

大きな χ^2 値をもつ分割表を2つ例示しよう。H_0 の下での期待度数を（　）内に示す。

	A_0	A_1	A_2
B_0	3 (3.616)	8 (11.571)	22 (25.313)
B_1	2 (11.571)	113 (122.223)	27 (48.455)
B_2	38 (25.313)	69 (48.455)	123 (108.482)

	E_0	E_1	E_2
B_0	3 (3.616)	22 (25.313)	8 (11.571)
B_1	24 (25.313)	74 (108.482)	44 (48.455)
B_2	20 (11.571)	48 (48.455)	162 (122.223)

(2)最小値の位置 i_1 によって分類すると、$i_1 = 1$ が低頻度で生じるため P 値は2.5～5%となるが、値との組 (i_1, t_{i_1}) に移ると、5%を越える結果になる。また最大値の位置 i_4 では $i_4 = 1$ が高頻度で生じ、P 値は0.1%未満になる。しかしながら、値との組 (i_4, t_{i_4}) に移ると5%を越えてしまう。最後に (i_1, i_4) によって分類すると、P 値は0.5～1%である。(2, 1)と(4, 1)が高頻度、(1, 4)と(4, 2)が低頻度を示す。

(3) $t_{(2)} = t_{(3)}$ となるケースに限って、(i_1, i_2, i_3, i_4) の12項目の分類を行うと、P 値が1～2.5%になるのが、H_0 を棄却できるケースである。

以上 t_1 の大きさに関する偏りが顕著で、いろいろな面から検出される。

芝(1) C に関する分割のみ P 値が2.5～5%になる。$t_2 < t_4$ が高頻度で生じる。

(2) i_4 による分類では H_0 を棄却できぬが、組 (i_4, t_{i_4}) に移って最大値によって細分すると、P 値が1～2.5%になる。これはダートの場合と反対の主張であり、いささか驚かされる結果である。また、$t_{(1)} = t_{(2)} < t_{(3)}$ となるケースで (i_1, i_2) によって分類すると、P 値は2.5～5%になるが、値との組 (i_1, i_2, t_{i_1}) に移ると5%より大きくなる。同じケースで (i_1, i_2, i_3, i_4) の12項目に細分すれば、P 値は1～2.5%になる。このとき、(2, 3, 1, 4)と(1 ,2, 4, 3)が高頻度、(2, 4, 3, 1)と(3, 4, 1, 2)が低頻度を示す。

以上芝の場合は、ダートと違って2, 3の面で偏りを示すに過ぎぬ。

最後にダートと芝を合わせたとき、χ^2 値はどう変化するか、目につく点を列挙する。C と E に関しては χ^2 値は分けた場合に比べて大きくなり、P 値は0.5%～1%と0.1%未満を示す。$A_i \cap E_j$, $C_i \cap D_j$, $C_i \cap E_j$, $D_i \cap E_j$, $E_i \cap F_j$ に移れば χ^2 値はやはり大きくなり、それぞれ0.1%未満, 1～2.5%, 0.1～0.5%, 0.5～1%, 0.1～0.5%の P 値を得る。他の場合はダートの大きな χ^2 値と芝の小さな χ^2 値の中間におさまる。合わせたときダートと違って H_0 を棄却できなくなるケースは、D と $D_i \cap F_j$ の2つ。次に、i_1 と (i_1, i_4) については合わせると χ^2 値が大きくなり、1～2.5%と0.5～1%の P 値を示す。i_4 の方は両者の中間に来て0.1～0.5%の P 値を得る。さらに組 (i_1, i_2) と (i_4, t_{i_4}) では χ^2 値は小さくなり、P 値は5%を越えてしまう。このようにダートと芝では相反する要素が含まれていて、明確な方向づけをすることは難しい。

(ロ)阪神競馬場：ダートの $n_1 = 473$ と芝の $n_2 = 250$ のレースを分析する。全レース $(N = 2484)$ 中に占める割合は、19.0%と10.1%である。

ダート(1) C と F において P 値は0.1～0.5%と1～2.5%になる。t_2, t_3 に比べて t_4 が大きい傾向を示す。偏差の大きい C を含む積事象を調べると、すべて H_0 が棄却される。$B_i \cap C_j$, $C_i \cap D_j$, $C_i \cap E_j$ で0.5～1%、$C_i \cap F_j$ (および $D_i \cap F_j$ と $E_i \cap F_j$)で1～2.5%、$A_i \cap C_j$ (および $B_i \cap E_j$)で2.5～5%の P 値を得る。

(2)最大値の位置について、$i_4 = 4$ が高頻度、$i_4 = 2$ が低頻度で生じ、0.1%未満の P 値を得るが、(i_4, t_{i_4}) に移ると、5%より大きくなる。(i_1, i_4) による分類では0.1～0.5%の P 値になる。(1, 4)と

(3, 4)が多く、(1, 2)と(1, 3)が少ない。

(3) t_i がすべて異なるケースで (i_1, i_2, i_3, i_4) の24項目で分類すると、0.5～1%の P 値を示す。(1, 2, 3, 4)と(3, 2, 1, 4)が高頻度、(1, 3, 4, 2)が低頻度である。

以上 t_4 の大きさに関する偏りが種々の面で見出される。しかしながら、次に論じる芝の場合が、ダートと同じ向きの圧倒的に大きな偏差をわれわれに呈示するので、上記の記述も色褪せてみえるかもしれぬ。

芝(1) A, C, D, F の分割において P 値はそれぞれ0.5～1%, 0.1%未満, 0.1%未満, 2.5～5%になる。t_2 が小さくて t_4 が大きい傾向が明白に(次の(2)でも同様に)読み取れる。C か D を含む積事象を調べると、0.5～1%の $B_i \cap D_j$ と $D_i \cap F_j$ を除く残り7つ(および $A_i \cap F_j$ と $E_i \cap F_j$)において、すべて0.1%未満の P 値を示すという驚嘆すべき結果を得る。なお、$A_i \cap E_j$ は2.5～5%の P 値である。

χ^2 値が極めて大きい分割表を3つ例示しよう。

	D_0	D_1	D_2
A_0	0 (0)	18 (12.5)	16 (12.5)
A_1	16 (12.5)	81 (50)	30 (50)
A_2	7 (12.5)	45 (50)	37 (50)

	D_0	D_1	D_2
C_0	1 (2.232)	5 (7.143)	7 (15.625)
C_1	11 (7.143)	115 (75.446)	27 (29.911)
C_2	11 (15.625)	24 (29.911)	49 (66.964)

	D_0	D_1	D_2
E_0	1 (2.232)	25 (15.625)	8 (7.143)
E_1	11 (15.625)	62 (66.964)	25 (29.911)
E_2	11 (7.143)	57 (29.911)	50 (75.446)

(2)最小値の位置 i_1 による分類では、$i_1 = 2$ の度数が大きく0.1%未満の P 値になり、組 (i_1, t_{i_1}) に移っても1〜2.5%の P 値を得る。また最大値の方は $i_4 = 4$ の度数が大きく0.1%未満の P 値、組 (i_4, t_{i_4}) に移っても0.1%未満の P 値が保持される。(4, 8)と(4, 9)が極めて高い頻度で生じる。さらに、(i_1, i_4) による分類でも P 値は0.1%未満であり、$t_{(1)} = t_{(2)} < t_{(3)}$ のケースでの (i_1, i_2) に移ると P 値は2.5〜5%である。

(3) t_i がすべて異なるケースで (i_1, i_2, i_3, i_4) の24項目で分類すると、P 値は0.1%未満になる。(2, 1, 3, 4)と(2, 3, 1, 4)が極めて高い頻度で生じる。

最後にダートと芝を合わせるとどうなるか、調べると多くの分類項目において χ^2 値は大きくなる。例えば、(i_1, i_2, i_3, i_4) に基づく分類(上記 t_i がすべて異なるパターンだけでなく、2つが一致するパターンにおいても χ^2 値は大きくなる)、$i_4, (i_4, t_{i_4}), (i_1, i_4)$ による分類、そして C, E (単独では両者5%を少し超えるが、合わせると5%未満になるケース)および F による分割、$A_i \cap E_j$, $B_i \cap C_j$, $B_i \cap E_j$, $B_i \cap F_j$ (E と同じく、合わせて初めて5%未満になるケース)、$C_i \cap D_j$, $C_i \cap F_j$, $D_i \cap F_j$ を挙げることができる。他の項目では合わせると、芝の大きい χ^2 値とダートの小さい χ^2 値の中間に来るけれども、そのうち P 値が0.1%未満になるのは $A_i \cap C_j$, $A_i \cap F_j$, $C_i \cap E_j$, $D_i \cap E_j$, $E_i \cap F_j$ の5つ。D では1〜2.5%, $A_i \cap D_j$ と $B_i \cap D_j$ では2.5〜5%の P 値を得る。

(ハ)京都競馬場：ダートの $n_1 = 468$ と芝の $n_2 = 167$ レースを分析する。全レース $(N = 2588)$ 中、18.1%と6.5%である。

ダート(1)A〜F 6種類の分割すべてにおいて、H_0 を棄却できる。P 値は阪神芝の場合ほど小さくないが、0.5〜1%が B と D、1〜2.5%が A, C, F、2.5〜5%が E である。t_3, t_4 が t_1, t_2 に比べて大きい傾向を示す。また、$t_1 = t_2$ が有意に多く、$t_3 = t_4$ が有意に少ない。次に積事象を調べると、5%をほんのわずか越える $B_i \cap D_j$ (不思議なことに、小さな χ^2 値をもつ芝と合わせると χ^2 値が大きくなって1〜2.5%の P 値を示す)を唯一つの例外として、残りすべてにおいて H_0 が棄却される。このうち最大の χ^2 値をもつ分割表を例示する。

	E_0	E_1	E_2
B_0	10 (4.179)	29 (29.25)	17 (13.371)
B_1	44 (29.25)	119 (125.357)	71 (55.993)
B_2	8 (13.371)	41 (55.993)	129 (141.236)

(2)最大値の位置 i_4 による分類で、 $i_4 = 4$ が高頻度で生じて1～2.5％の P 値を得るが、 組(i_4, t_{i_4}) に移ると5％よりかなり大きくなる。また、 $t_{(1)} = t_{(2)} < t_{(3)}$ のケースで (i_1, i_2) に着目すると、(1, 2)と(2, 3)が高頻度、(3, 4)が低頻度で生じ、1～2.5％の P 値を示す。値との組 (i_1, i_2, t_{i_1}) に移っても1～2.5％の P 値が保持される。このとき、$(i_1, i_2) = (1, 2)$ に対しては $t_{i_1} \geqq 2$ の方が多く、(2, 3)に対しては $t_{i_1} = 1$ の方が多く現れる。

(3) $t_{(1)} = t_{(2)} < t_{(3)}$ のケースで、 (i_1, i_2, i_3, i_4) の12項目で分類すると、上記 (i_1, i_2) による場合と同じく1～2.5％の P 値を得る。少数例ながらも、 t_i のうち3つの値が一致するケースでは、値による分類でも位置による分類でも、両方とも P 値は5％より小さくなることを付言する。

芝　すべての項目において H_0 を採択する結果に終わる。

　最後に両者合わせると、ダートで H_0 が棄却されたところは、そのまま H_0 が棄却される。ダートの場合からの P 値の変動はそう大きくはない。つまり、芝のレース結果が、ダートの偏りを打ち消す方向に働くことはほとんどみられない。

(ニ)中山競馬場：ダートは $n_1 = 845$ 、芝は $n_2 = 289$ で $m = 16$ 頭のレースが最も多く行われる。全レース $(N = 2663)$ 中、31.7％と10.9％を占める。

ダート　ほとんどの分類項目で H_0 が採択されるが、例外は次の5つ。

(1) $t_1 > t_2$ の方に偏る E の分割、および $A_i \cap C_j$, $C_i \cap F_j$ の3つにおいて2.5～5％の P 値になる。

(2)最小値の位置 i_1 では P 値は5％をはるかに越えるけれども、値との組 (i_1, t_{i_1}) に移ると、0.5～1％の P 値になる。 $t_{i_1} = 1$ のとき $i_1 = 1$ が少なく、 $i_1 = 4$ が多いのに対照的に、 $t_{i_1} = 2$ のときは $i_1 = 1$ が多く、 $i_1 = 4$ が少ない。

(3) t_i のうち3つが一致するパターン(度数は34)で、他と異なるのが t_1 であるケースが半分近くも占める。(この結果、上記 $A_i \cap C_j$ と $C_i \cap F_j$ における χ^2 値が大きくなる。)

芝(1) t_1 が小さく、 t_4 が大きい傾向を示す。実際 B と D による分割で、ともに P 値は1～2.5％になる。積事象ではすべて H_0 を採択する結果になる。

(2)最大値の位置 i_4 で分類すると、2.5～5％の P 値を得る。最小値の方は、単独の i_1 では5％を少し越えるが、値との組 (i_1, t_{i_1}) に移れば1～2.5％の P 値になる。 $t_{i_1} = 1$ のときダートの場合と反対に $i_1 = 1$ が多く、 $i_1 = 4$ が少ない。 $t_{i_1} \geqq 2$ のときは、 $i_1 = 2$ が少ない。従って芝とダートを合わせると、 P 値は5％より大きくなる。最後に、 (i_1, i_4) において2.5～5％の P 値を得る。

　ダートと芝を合わせると、互いに打ち消し合う方向に働くため H_0 を棄却できる分類項目を発見することができない。これは[3]の結論に合致する。

(ホ)東京競馬場：ダートは $n_1 = 503$ 、芝は $n_2 = 126$ で、全レース $(N = 2483)$ 中20.3％と5.1％を占める。 $m = 16$ 頭による芝のレースの少なさが際立つ。

ダート(1) D による分類のみ2.5～5％の P 値であり、 $t_1 < t_4$ の向きに偏る。積事象では $B_i \cap F_j$

（$t_1 < t_3 < t_4$ の度数が大きい）, $C_i \cap D_j$, $D_i \cap E_j$ において、それぞれ0.1～0.5%, 1～2.5%, 2.5～5%の P 値を得る。

(2)最大値の位置に関して、$i_4 = 4$ が高頻度で生じて2.5～5%の P 値になるが、値との組 (i_4, t_{i_4}) に移ると5%より大きくなる。

(3)特筆すべきは、5つのタイプ(no pair型、3つのone pair型、three cards型)への分類において0.5～1%の P 値が得られること。他の競馬場ではみられぬ特色である。$t_{(1)} = t_{(2)}$ のケースが多く、$t_{(3)} = t_{(4)}$ のケースが少ない。また高頻度で生じる $t_{(1)} = t_{(2)}$ のケースに限ると、(i_1, i_2, i_3, i_4) による12項目において(1, 2, 3, 4)と(1, 3, 2, 4)が高頻度で生じ、P 値は2.5%～5%になる。しかしながら、最初の (i_1, i_2) だけみて分類すれば、5%をはるかに越えてしまう。

芝　H_0 を棄却できる分類項目は数少ない。まず C による分割で、$t_2 > t_4$ のケースが有意に少ない。最小値の位置では $i_1 = 3$ の度数が最大で、2.5～5%の P 値になる。ダートの場合は $i_1 = 3$ が最小度数であり、両者合わせると χ^2 値が小さくなって5%より大きくなる。最後に、特異な分割表を示すものとして (i_4, t_{i_4}) を取りあげよう。$i_4 = 4$ の行では、$t_4 \geqq 11$ が極めて大きい度数をもち、$i_4 = 3$ の行では $t_3 = 7$ の度数が目立つ。(阪神芝では、$i_4 = 4$ で $t_4 = 8, 9$ の度数が極端に大きかった点を思い起こす。)また $t_{i_4} = 10$ の度数が期待度数よりも小さい。例数が少ないけれども、χ^2 値を計算して1～2.5%の P 値を導く。

さて、ダートと芝を合わせた場合、χ^2 値が大きくなり5%未満の P 値を示すのは、C による分割、$B_i \cap C_j$ と $C_i \cap D_j$, そして $t_{(1)} = t_{(2)}$ のケースで (i_1, i_2, i_3, i_4) による分類の4つ。なお、$B_i \cap F_j$, i_4 と (i_4, t_{i_4}) による分類では、合わせると両者の χ^2 値の中間に来て、それぞれ1～2.5%、2.5～5%と1～2.5%の P 値を与える。

謝　辞

資料の整理と原稿の清書をお願いした鴨藤江利子さんに、心から御礼申しあげます。また鋭いコメントで不十分な点を指摘し、著者の目を新たな方向に向けてくれた[2][3]のレフェリー達に感謝申しあげます。

参考文献

[1] Everitt BS : *The Analysis of Contingency Tables*, 2nd edition. Boca Raton : Chapman&Hall/CRC,1992.

[2] 野田明男：競馬データにみられる統計的偏りについて(1), 浜松医科大学紀要一般教育18: 1-11, 2004.

[3] 野田明男：競馬データにみられる統計的偏りについて(2), 浜松医科大学紀要一般教育19: 1-7, 2005.

[4] レーシングファイル(中央競馬全レース成績書), No.22～42. ケイバブック, 1999～2004.

[5] Peccati G : Hoeffding-ANOVA decompositions for symmetric statistics of exchangeable observations. *Ann.Probab*. 32(3A) : 1796-1829, 2004.