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　第 18 号（2004）

競馬データにみられる統計的偏りについて（1）

野田　明男
（数学）

On the Statistical Bias Found in the Horse Racing Data (1)

Akio NODA
Mathematics

Abstract: The purpose of the present paper is to report what type of statistical bias the author has found in the horse racing data (based on [3]).

In order to explain the type of statistical bias, let us consider a racing with m participants. We denote by $\{a, b, c\}(1 \leqq a < b < c \leqq m)$ a set of numbers of the first, second and third racehorses to reach the goal. The number of each participant is determined by lot, which leads us to the following null hypothesis:
H_0: A set {a,b,c} is nothing but a result of random sampling from the set {1,2,⋯,m}.
Studying the probability distributions of various random variables arising in the random sampling of H_0, we are in a position to examine, by means of the chi-square test, how frequency distributions observed in the data mentioned above deviate from the expected ones under the null hypothesis H_0.

Our method of contracting the original data consists of studying two random variables, $R = c - a$ (the range) and $D = \min\{b - a, c - b\}$ (the adjacent interval of three numbers), as well as the following pair of partitions of the total event:

$$A_0 = \{2b = a + c\},\ A_1 = \{2b < a + c\} \text{ and } A_2=\{2b > a + c\};$$
$$B_0 = \{a + b = c\},\ B_1 = \{a + b < c\} \text{ and } B_2 = \{a + b > c\}.$$

In this paper (1), we take up three racetracks, Chukyo, Hanshin and Kyoto, to examine all racings of $m = 16$ (and also $m = 14$) carried out on these racetracks. Indeed, we sum up the original data into two kinds of contingency tables, the one corresponding to the joint probability distribution of (R,D) and the other to the 3×3 probability table of the product events $A_i \cap B_j$ $(i, j = 0, 1, 2)$. Performing the chi-square tests for these contingency tables, we are able to detect some types of statistical bias for each racetrack. Furthermore, these results tell us interesting dependency of the type of detected bias upon the racetrack, which suggests that the individual character of racetrack can be extracted from the long-term racing files [3].

Keywords: contingency table, chi-square test, random sampling as the null hypothesis, adjacent interval of three numbers.

§1. 序 ： 帰無仮説とその下での確率分布

競馬レースに興味を覚えて約5年が経ち、その間の成績データ([3]参照)の中に潜む知の一つとして、統計的な偏りを発見する；見出された偏りのパターンをレースが行われる競馬場間で比較し、有意差を検出すれば競馬場の｢個性｣が見えて来る。著者が表題の論文シリーズで実践して行きたいと思うデータ解析の目標は、データから如何にして個性を抽出するか、という研究課題の解明である。

競馬レースの成績表はさまざまな面を合わせもち、情報量の大きなデータであるが、3 連複の馬券方式が導入されて競馬ファンに歓迎されたことに示唆されて、われわれは 1、2、3 着に入った馬番だけに目をつける：m 頭出走のレースにおいて、上位3頭の組合せ$\{a, b, c\}(1 \leqq a < b < c \leqq m)$はどのような頻度分布を示すのか。この問題に限定してデータを縮約し、集計作業を行う。

さて、1からmまでの馬番は、公正なくじによって出走馬に割り当てられる点を考慮すれば、次の帰無仮説H_0が、頻度分布における偏りを見出すための自然な基準を形作ることがわかる(医学統計学でよく出会う帰無仮説とその検定は、丹後俊郎氏の著作(例えば[2])から学ぶことができる)。

H_0：1からmまでの番号の中から、$\{a, b, c\}$はランダムに選ばれる。

ここで、$\{a, b, c\}$の組合せの総数は${}_mC_3 = \dfrac{m(m-1)(m-2)}{6}$であり、同程度に確からしいと考えるのが$H_0$に他ならぬ。

偏りのパターンを発見するための予備調査の結果から、上位3頭の馬番間の隣接の度合いに着目する：

$R = c - a$(範囲の幅)、$D = \min\{b-a, c-b\}$(隣接間隔)。

また、Dの最小値の決まり方に応じて、3つの事象に分割する：

$$A_0 = \{2b = a + c\}, A_1 = \{2b < a + c\}, A_2 = \{2b > a + c\}。$$

さらに、これらの事象と1対1対応がつけられる事象

$$B_0 = \{a + b = c\}, B_1 = \{a + b < c\}, B_2 = \{a + b > c\}$$

も合わせて考察する。上位3頭の馬番間の差によって定まる確率変数 R, D と事象A_iに対比して、事象B_jは内枠・外枠の差異に敏感な面を有することに注意しよう。すなわち、内枠の2頭が上位3頭の中に入ればB_1が起こりやすく(京都競馬場の特色の1つ)、他方外枠の2頭が入ればB_2が起こりやすい(中京競馬場の特色の1つ)。帰無仮説 H_0 の下での(R, D)の同時確率分布と、積事象$A_i \cap B_j$ $(i, j = 0, 1, 2)$ の確率表は、この節の後半で記述する。

この論文(1)では、中京、阪神、京都、3つの競馬場における約5年間のレース成績を分析する。出走頭数 m はレース毎に変化するが、最多レースのm=16 に焦点をあてる；[3]に記載されたレースの総数は、(イ)中京ではN_1=1428、(ロ)阪神ではN_2= 2292、(ハ)京都ではN_3= 2396である。このうち、m=16 のレース(出走取消の入った少数のレースは除く)の回数は、(イ) n_1=616(全体の中の43.14％を占める)、(ロ) n_2=656(28.62％)、(ハ) n_3=590(24.62％)となっている。このよう

な n_k 個のレース結果 $\{a,b,c\}$ を各 k $(k = 1, 2, 3)$ 毎に集計し、$A_i \cap B_j$ の確率表に対応する分割表Ⅰと(R,D)の確率分布表に対応する分割表Ⅱをそれぞれ作成する（§2[Ⅰ][Ⅱ]を参照）。得られた分割表に χ^2 検定を実行し、H_0を棄却できるかどうか調べる。H_0 からの偏りを見出した後さらに、3つの競馬場間で有意差検定を実行し、偏りの出方に差異があるかどうか判定する。次節で詳述されるこのような偏りと差異は、競馬場の個性が長期間にわたるレース結果に反映したものである、と受け取ることができよう。

[3]に記載されているレースはすべて、$6 \leqq m \leqq 18$ の範囲に入るが、$m = 16$ に次いで多いのは $12 \leqq m \leqq 15$ の範囲内の m の値である。出走取消によって欠番が生じたレースはわれわれの集計作業から除外する。§2の最後の部分[Ⅲ]で $m = 14$の場合（レースの回数は、(イ)中京では $n_1' = 140$（総数 N_1で割れば9.80%)、(ロ)阪神では $n_2' = 267$(11.65%)、(ハ)京都では $n_3' = 225$(9.39%)となっている）だけ検定結果を報告し、前記の $m = 16$ の場合と対比させる。$\{a,b,c\}$ の頻度分布は m の値が変わるとどのように反応するのか、という難問および、東京、中山、新潟など他の競馬場においてはどんな偏りパターンがわれわれを待ちうけているのか、という課題は、データに潜む知を引き出すための新しい視点（レフェリーから指摘された「内枠・外枠の問題」に切り込む方法など）を模索しつつ、同一表題の(2)として取り組む計画をたてている。

この節の残りは、帰無仮説 H_0 に基づく確率計算の結果を示す（証明は容易であり、省略する）。次節の χ^2 検定において、期待度数を算出するのに必要である。正の実数 a に対し、$\lfloor a \rfloor$ は小数点以下切り捨てた整数、$\lceil a \rceil$ は切り上げた整数を表す。

<u>命題1.</u>　$2 \leqq r \leqq m-1,\ 1 \leqq d \leqq \left\lfloor \frac{m-1}{2} \right\rfloor$に対し、$R = r$ かつ $D = d$ となる場合の数 x は、次式で与えられる。

(1) r が奇数のとき：$d \leqq \frac{r-1}{2}$ なら $x = 2\,(m-r)$;　$d \geqq \frac{r+1}{2}$ なら $x = 0$。

(2) r が偶数のとき：$d \leqq \frac{r}{2}-1$ なら $x = 2\,(m-r)$;　$d = \frac{r}{2}$ なら $x = m-r$;
$d \geqq \frac{r}{2}+1$ なら $x = 0$。

$m = 16$ の場合に、(R,D)の H_0 の下での同時確率分布表を記す（すべての場合の数560で割算する形）。

Rの値	2	3	4	5	6	7	8	9
$D = 1$	14	26	24	22	20	18	16	14
$D = 2$	0	0	12	22	20	18	16	14
$D = 3$	0	0	0	0	10	18	16	14
$D = 4$	0	0	0	0	0	0	8	14
計	14	26	36	44	50	54	56	56

Rの値	10	11	12	13	14と15	計
$D=1$	12	10	8	6	6	196
$D=2$	12	10	8	6	6	144
$D=3$	12	10	8	6	6	100
$D=4$	12	10	8	6	6	64
$D\geqq 5$	6	10	12	12	16	56
計	54	50	44	36	40	560

命題2. (R,D)の周辺分布は次の通り：

$$P(R=r)=\frac{6(r-1)(m-r)}{m(m-1)(m-2)}\quad (2\leqq r\leqq m-1)$$

$$P(D=d)=\frac{6\,(m-2d)^2}{m(m-1)(m-2)}\quad \left(1\leqq d\leqq \left\lfloor\frac{m-1}{2}\right\rfloor\right)$$

積事象 $A_i\cap B_j$ の場合の数をかぞえるため、まず単純な1対1対応 $(a,b,c)\mapsto(\alpha,\beta,\gamma)$ を4種類構成する。整数 a に対し、a を2で割った余りを a' と書く。

（イ） $\alpha=m+1-c,\quad \beta=m+1-b,\quad \gamma=m+1-a$ によって定まる全事象上の1対1対応 f は A_0 を不変にし、A_1 を A_2 に、A_2 を A_1 に写す。

（ロ） $\alpha=\dfrac{a+a'}{2},\quad \beta=2b-\dfrac{3a+a'}{2},\quad \gamma=c$ によって定まる和事象 $A_0\cup A_1$ 上の1対1対応 g_1 は、A_0 を B_0 に、A_1 を B_1 に写す。

（ハ） $\alpha=\left\lceil\dfrac{c-a}{2}\right\rceil,\quad \beta=b,\quad \gamma=c$ によって定まる $A_0\cup A_2$ 上の1対1対応 g_2 は、A_0 を B_0 に、A_2 を B_2 に写す。

（ニ） $\alpha=b-a,\quad \beta=b,\quad \gamma=c$ によって定まる $A_0\cap(B_1\cup B_2)$ 上の1対1対応 h は、$A_0\cap B_1$ を $B_0\cap A_1$ に、$A_0\cap B_2$ を $B_0\cap A_2$ に写す。

（イ）（ロ）（ハ）の対応によって、 $P(A_0)=P(B_0)$ と $P(A_1)=P(A_2)=P(B_1)=P(B_2)$ が導かれ、前者の確率 $P(A_0)=P(B_0)=p_m$ とおくと、後者の確率はすべて $\dfrac{1-p_m}{2}$ となる。

命題3.

$$p_m=\frac{6\left\lfloor\frac{m}{2}\right\rfloor\left\lceil\frac{m-2}{2}\right\rceil}{m(m-1)(m-2)}$$ が成り立つ。

最後の対応（ニ）を用いると、$P(A_0\cap B_1)=P(B_0\cap A_1), P(A_0\cap B_2)=P(B_0\cap A_2)$ が得られ、$P(A_1\cap B_2)=P(A_2\cap B_1)$ が従う。こうして $A_i\cap B_j\,(i,j=0,1,2)$ の確率表を作成するには、3つの積事象 $A_0\cap B_0, A_0\cap B_1, A_2\cap B_2$ に関する場合の数 y,z,w を求めればよい。なお、表記の簡便化から、$\langle a\rangle^2=a(a+1), \langle a\rangle^3=a(a+\frac{1}{2})(a+1)$ と書く。

命題4. 次式が成り立つ。

（1）$y=\left\lfloor\frac{m}{3}\right\rfloor$

（2）$z=\frac{1}{2}\left\langle\left\lfloor\frac{m-2}{3}\right\rfloor\right\rangle^2+(m-3)\left(\left\lfloor\frac{m-3}{2}\right\rfloor-\left\lfloor\frac{m-2}{3}\right\rfloor\right)-\left\lfloor\frac{m-3}{2}\right\rfloor^2+\left\lfloor\frac{m-2}{3}\right\rfloor^2$

（3）
$$\begin{aligned}w=\frac{1}{2}\Bigg[&m\left\langle\left\lfloor\frac{m}{3}\right\rfloor-1\right\rangle^2+\left(m-\frac{3}{2}\right)\left\langle\left\lfloor\frac{m-2}{3}\right\rfloor\right\rangle^2\\&+\left(\left\lfloor\frac{m}{2}\right\rfloor-\left\lfloor\frac{m}{3}\right\rfloor-1\right)\left\{\langle m-1\rangle^2-(2m-1)\left(\left\lfloor\frac{m}{2}\right\rfloor+\left\lfloor\frac{m}{3}\right\rfloor\right)\right\}\\&+\left(\left\lfloor\frac{m-3}{2}\right\rfloor-\left\lfloor\frac{m-2}{3}\right\rfloor\right)\left\{\langle m-2\rangle^2-(2m-3)\left(\left\lfloor\frac{m-3}{2}\right\rfloor+\left\lfloor\frac{m-2}{3}\right\rfloor+1\right)\right\}\Bigg]\\&+\frac{2}{3}\left\{\left\langle\left\lfloor\frac{m}{2}\right\rfloor-1\right\rangle^3+\left\langle\left\lfloor\frac{m-3}{2}\right\rfloor\right\rangle^3-\left\langle\frac{m}{3}\right\rangle^3-\left\langle\left\lfloor\frac{m-2}{3}\right\rfloor\right\rangle^3\right\}\\&-\frac{5}{6}\left\{\left\lfloor\frac{m}{3}\right\rfloor^3-\left\lfloor\frac{m}{3}\right\rfloor+\left\langle\left\lfloor\frac{m-2}{3}\right\rfloor\right\rangle^3\right\}\end{aligned}$$

$m=16$ の場合に、$A_i\cap B_j$ の場合の数を3×3の表の形に示して、この節を終える。

	B_0	B_1	B_2	計
A_0	5	16	35	56
A_1	16	169	67	252
A_2	35	67	150	252
計	56	252	252	560

§ 2. 中京・阪神・京都競馬場でのレース成績に対する χ^2 検定

この節では、(イ)中京(ロ)阪神(ハ)京都、3つの競馬場における $m=16$ (および14)のレース成績を取り上げて、前節で述べた視点から χ^2 検定を実行し、その結果を詳述する。

初めに、$m=16$ のレース結果 $\{a,b,c\}$ を、取り扱いやすさから[Ⅰ]積事象 $A_i \cap B_j$へ3×3の形に分類して、集計した分割表Ⅰを提示し、これに対して帰無仮説 H_0 が棄却できるかどうか調べる。さらに、3つの競馬場間の有意差検定を実行する。次に同じ $m=16$ のデータを、[Ⅱ](R,D)の同時確率分布表に対応した形式に集計した分割表Ⅱを作成し、これに対して χ^2 検定をいろいろな視点から実行する。

最後の[Ⅲ]の部分は、$m=14$ の場合にあてられる；上記[Ⅰ][Ⅱ]と対比させる形で、検定結果をそれぞれ簡潔に述べて行く。他の m の値および他の競馬場のレース成績については、次の論文(2)において分析して行く予定である。

[Ⅰ]出走取消により欠番が生じたレースは除外して、$m=16$ のレース結果を[3]により集計し、次の分割表Ⅰに至る。

Ⅰ	B_0	B_1	B_2	計
A_0	5	6	60	71
	12	22	41	75
	11	16	39	66
A_1	19	188	94	301
	21	213	79	313
	19	210	61	290
A_2	28	45	171	244
	43	58	167	268
	35	72	127	234
計	52	239	325	616
	76	293	287	656
	65	298	227	590

上段の数字は(イ)における度数

中段の数字は(ロ)における度数

下段の数字は(ハ)における度数

を示す。以下同じ方式を採用する。

前節の H_0 の下での確率表から期待度数を各欄で求めて、χ^2 統計量を導く（流布している χ^2 分布表に従って、小数第4位まで出す）。計算結果は

（イ）39.6820　（ロ）14.3014　（ハ）19.9286

となる。この値への寄与は、（イ）では対角線上にある $A_i \cap B_i$ $(i=0,1,2)$ と $A_1 \cap B_0$ 以外のすべての欄で大きく、反対に（ハ）では対角線上の3つの欄での距たりが大きい。（ロ）では $A_0 \cap B_0$ が起こりやすく、$A_2 \cap B_1$ が起こりにくい点が目につくけれども、自由度 $\nu=8$ で調べると、P 値は5％と10％の間に入り、H_0を棄却できぬ結果に終わる。

<u>結論Ⅰ-1.</u>　$A_i \cap B_j$ $(i, j = 0,1,2)$ の3×3の形に分類するとき、

（イ）中京競馬場では、有意水準0.1％で帰無仮説 H_0を棄却する。

（ロ）阪神競馬場では、有意水準5％で帰無仮説 H_0を採択する。

（ハ）京都競馬場では、有意水準2.5％で帰無仮説 H_0を棄却する。

次に、偏りのパターンが（イ）と（ハ）であざやかな対照をなす事実に留意して、3群間の有意差検定（一様性検定）を行う。つまり、$A_i \cap B_j$ の起こる確率は、その欄の3つの度数の和とレースの総数1862との比で推定する。このとき、χ^2=40.5315 と計算される。ν=16 なので、検定結果は次のようにまとめられ、われわれの目的が達成される。

<u>結論Ⅰ-2.</u>　（イ）（ロ）（ハ）3群間に有意水準0.1％で有意差が認められる。

［Ⅱ］m=16 のレースを［3］に基づき集計して、次の分割表Ⅱを得る。

Ⅱ	R=2	R=3	R=4	R=5	R=6	R=7	R=8	R=9
D=1	<u>24</u> <u>26</u> <u>27</u>	<u>35</u> <u>34</u> <u>34</u>	<u>23</u> <u>32</u> <u>25</u>	<u>21</u> <u>30</u> <u>17</u>	<u>29</u> <u>23</u> <u>30</u>	<u>19</u> <u>31</u> <u>19</u>	<u>19</u> <u>18</u> <u>10</u>	<u>24</u> <u>16</u> <u>18</u>
D=2			<u>25</u> <u>20</u> <u>15</u>	<u>30</u> <u>39</u> <u>25</u>	<u>22</u> <u>20</u> <u>19</u>	<u>21</u> <u>34</u> <u>18</u>	<u>21</u> <u>16</u> <u>17</u>	<u>16</u> <u>14</u> <u>15</u>
D=3					<u>7</u> <u>6</u> <u>7</u>	<u>23</u> <u>24</u> <u>17</u>	<u>15</u> <u>19</u> <u>18</u>	<u>15</u> <u>18</u> <u>11</u>
D=4							<u>8</u> <u>7</u> <u>6</u>	<u>13</u> <u>16</u> <u>12</u>
計	<u>24</u> <u>26</u> <u>27</u>	<u>35</u> <u>34</u> <u>34</u>	<u>48</u> <u>52</u> <u>40</u>	<u>51</u> <u>69</u> <u>42</u>	<u>58</u> <u>49</u> <u>56</u>	<u>63</u> <u>89</u> <u>54</u>	<u>63</u> <u>60</u> <u>51</u>	<u>68</u> <u>64</u> <u>56</u>

つづき	R=10	R=11	R=12	R=13	R≧14	計
D=1	8	11	17	8	5	243
	15	10	7	6	6	254
	12	10	9	4	5	220
D=2	10	16	8	5	5	179
	12	9	7	5	6	182
	13	4	9	7	10	152
D=3	12	8	5	4	1	90
	12	11	5	5	5	105
	13	7	14	5	4	96
D=4	13	10	4	7	7	62
	17	9	9	2	8	68
	13	15	5	6	7	64
D≧5	5	9	6	8	14	42
	6	7	13	10	11	47
	6	12	12	11	17	58
計	48	54	40	32	32	616
	62	46	41	28	36	656
	57	48	49	33	43	590

2つの確率変数RとDをともに考える今の場合、次の3つの視点から分割表Ⅱを分析しよう。
1. Dの値による分類　2. Rの値による分類　3. (R,D) 両者の値を対にした45の欄への分類

［Ⅱ-1］われわれは $D≧5$ を1項目にしたため、自由度は $v=4$ である。χ^2統計量を計算すると、
（イ）17.0366　（ロ）10.8247　（ハ）1.9054を得る。

<u>結論Ⅱ-1.</u>　Dの値によって5つに分類するとき、
（イ）中京競馬場では、有意水準0.5％で、H_0を棄却する。
（ロ）阪神競馬場では、有意水準5％で、H_0を棄却する。
（ハ）京都競馬場では、有意水準5％で、H_0を採択する。
　いずれの競馬場においても、$D≦2$ の頻度が期待値よりも高く、$D≧3$ の頻度は低い；Dの値を2以下と3以上に2分割して、χ^2検定を実行すると、χ^2統計量はそれぞれ、
（イ）15.6811　（ロ）9.0904　（ハ）1.3505となる。

<u>結論Ⅱ-2.</u>　$D≦2$ と $D≧3$ との2項目に区分するとき、
（イ）中京競馬場では、有意水準0.1％で H_0を棄却する。
（ロ）阪神競馬場では、有意水準0.5 ％で H_0を棄却する。
（ハ）京都競馬場では、有意水準5％で H_0を採択する。

ところで、帰無仮説H_0は忘れて、3つの競馬場間の有意差検定を行うと、[Ⅰ]とは異なり、有意差は検出されない；Dの値に基づく偏りパターンが3つの競馬場でそろっているためと考えられる。

[Ⅱ-2] Rの値による分類では、14と15の値を1つにまとめたので、v=12 となり、χ^2統計量は次の通り。

(イ)17.6300　(ロ)38.9499　(ハ)15.1080

結論Ⅱ-3.　Rの値によって分類するとき、

(イ) 中京競馬場では、有意水準5％で H_0を採択する。

(ロ) 阪神競馬場では、有意水準0.1％で H_0を棄却する。

(ハ) 京都競馬場では、有意水準5％で H_0を採択する。

偏りが検出された(ロ)の場合、R≦5 の頻度は期待値よりも高く R≧8 の頻度は低い；特筆すべきはR=7 の度数が異常に大きいことであり、隣の6の度数は期待値より下がる。これに対比して、(イ)の場合はR≦9 が起こりやすく、R≧10 は起こりにくい；試みに、9以下と10以上とで2分してみると、χ^2=11.0400 となって、H_0を有意水準0.1％で棄却できることになる。また、(ハ)の場合R=2 の頻度が異常に高い。

次に(イ)(ロ)(ハ)共通の特徴として、R≦4 の度数が期待値より大きい点に着目して、Rの値を4と5の間で切断すれば、χ^2はそれぞれ

(イ)7.5782　(ロ)6.8579　(ハ)6.3292

と計算される。

結論Ⅱ-4.　R≦4 と R≧5 の2項目に区分するとき、

(イ) 中京競馬場では、有意水準1％で H_0を棄却する。

(ロ) 阪神競馬場では、有意水準1％で H_0を棄却する。

(ハ) 京都競馬場では、有意水準2.5％で H_0を棄却する。

ところで、Rの値に基づいて、3つの競馬場間の有意差検定を行うと、χ^2=23.8924 を得る。v=24 で、有意差は認められないという結果に終わる。

[Ⅱ-3] (R,D)両者の値を対にすると、分類項目は45にも上り、細分し過ぎの恐れがあるが、前節で求めた同時確率分布表から期待度数をそれぞれ求め、分割表Ⅱの χ^2統計量を導く。

(イ)62.9768　(ロ)52.9222　(ハ)42.5855

結論Ⅱ-5.　(R,D)両者の値に基づいて分類するとき、

(イ) 中京競馬場では、有意水準5％で H_0を棄却する。

（ロ）阪神競馬場では、有意水準5％で H_0 を採択する。

（ハ）京都競馬場では、有意水準5％で H_0 を採択する。

m=16のレース成績について［Ⅰ］［Ⅱ］の結論を要約する：（イ）の偏りは顕著であり、いろいろな方法で検出できる。（ロ）の偏りは、R（およびD）の値に基づくとうまく検出できる。（ハ）の偏りは、$A_i \cap B_j$ の3×3の表への分類が検出方法としてすぐれている。競馬場に適した検出方法を見出す必要がある。

［Ⅲ］m=14の場合は、16の場合と比べるとレース数が大幅に減少する：（イ）140（ロ）267（ハ）225。集計結果を［Ⅰ］［Ⅱ］のように分割表の形で提示することは省略し、χ^2検定結果を箇条書きに述べて行く。

1. $A_i \cap B_j$ による分類では、$A_0 \cap B_0$ の期待度数が小さく、$A_1 \cap B_0$（または $A_0 \cap B_1$）と合併するとν=7である。（イ）（ロ）に対しては有意水準5％で H_0 を棄却できないけれども、（ハ）については有意水準2.5％で H_0 を棄却する。$A_2 \cap B_2$ の度数が期待値よりも極めて小さい（m=16でもこの通り）点が目立つ。
2. Dの値による分類では、（イ）（ロ）（ハ）とも有意水準5％で H_0 を棄却できぬ。しかしながら、（ロ）では D=2の度数が異常に小さい点に着目する；D=2と $D \neq 2$ とに分けると、χ^2=6.3305を得、有意水準2.5％で H_0 を棄却する。
3. Rの値による分類では、（イ）（ロ）（ハ）とも χ^2 の値は小さく、H_0 を採択せざるを得ぬ。しかしながら、（ロ）において期待値からの偏差の符号を調べて、$R \leqq 3$, $4 \leqq R \leqq 5$, $6 \leqq R \leqq 9$, $10 \leqq R \leqq 11$, $R \geqq 12$, という風に5つに区分すると χ^2=9.8958を得て、有意水準5％で H_0 を棄却できる。
4. (R,D)両者の値を対にする分類では、レース数の少なさと分類項目の多さが災いして、めぼしい結果を導くのが容易でない。しかしながら、（イ）における不思議な偏りのパターンがR=4と5の部分に出現する：(R,D)が(4,1)と(5,2)の欄は期待値よりも度数が高く、(4,2)と(5,1)の欄は反対に度数が低い。残りを一つにまとめると（ν=2）、χ^2=13.9668を得、有意水準0.1％で H_0 を棄却する。なお、（イ）のm=16の場合には、(4,2)と(5,2)の欄で度数が高く、(4,1)と(5,1)の欄で度数が低くなっている。同様に残りを一まとめにすると、χ^2=9.3724を得、有意水準1％で H_0 を棄却する。

   $12 \leqq m \leqq 15$ の範囲のレース結果に関しては、m の値の変化に応じて偏りのパターンがどう変わるかに留意しつつ、総合的に分析しなければならぬ。ひき続く⑵でこの研究課題に取り組む計画である。

謝　辞

著者の競馬のおもしろさへの開眼は、研究室の隣接間隔Dが小さい佐藤弘明教授と旬の魚を食べさせてくれる町随一の店主金原慶彦氏、御二人の導きに負っている。また、競馬データの解析に本腰を入れるキッカケは、平成14年度の基礎配属で研究室に来てくれた新海宏明君と岩川恵美さんの

おかげである。

この夏、集計作業の一部を娘の紗希が手伝ってくれた。最後に、〆切りの迫る最中、手間のかかる表の作成を鴨藤江利子さんにお願いした。以上の方々に深謝する次第です。

参考文献

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

[2] 丹後俊郎：統計モデル入門. 朝倉書店, 2000.

[3] レーシングファイル（中央競馬全レース成績収録）, No.22～41. ケイバブック, 1999～2003.