筑波大学体育科学系紀要　Bull. Health & Sports Sciences, Univ. of Tsukuba 9：159—174，1986.

統計的仮説検定を用いた個体の分類と適当な空間における個体の布置の決定

—統計的仮説検定手続を導入したクラスター分析手法と個体の布置決定の工夫—

松　浦　義　行

A novel procedure of individual classification using statistaical inference and determination of individual configuration in a certain proper space.

—A new procedure of cluster analysis and determination of individual configuration in a proper space—

Yoshiyuki MATSUURA

In most procedures of cluster analysis, the number of clusters is determined arbitrarily without any objective criterion for judgement. This is not only one of their strong points but also one of their weak points. In the new procedure devised in this paper, statistical inference was utilized to judge whether one individual belongs to a certain cluster or not. For applying statistical inference; testing statistical hypothesis, the inter-individual distances must be proved to follow to a certain distribution whose probability density function is known. This is one of the great restrictions for application of this procedure, but the arbitrary judgement can be prevented. This procedure was originally devised to lead to the non-hierarchical clustering solution; however, it also can lead to the hierarchical clustering solution through repeating application of this procedure to the distance matrix whose elements are the inter-cluster distances. Therefore the definition of distance between individual and cluster, and inter-cluster distance was discussed in this paper.

Then, a new estimation procedure of inner-product matrix from the distance matrix given was devised, although this idea was basically similar to that of Young and Housholder (1938).

These ideas were applied to the calssification of twelve college sports teams, and it was shown that the hierarchical clustering could be derived through applying non-hierarchical clustering porocedure repeatedly. Then, the configuration of these twelve sports teams was evaluated in a certain proper two dimensional orthogonal spaces derived from the distance matrix with three different procedures.

序　論

クラスター分析は，$x_1x_2x_3\cdots\cdots x_n$のｎ個の個体が与えられた時，これらｎ個の個体の類似性，または非類似性を手がかりとして，ｎこの個体をｍ（$m \leqq n$）のクラスターに分類することを目的とした統計的方法の一種である。

類似性，非類似性をあらわす統計値には種々のものが工夫されているが，類似性には相関係数，積和等が，非類似性には個体間の距離が広く用いられている。この距離をあらわす統計値にも種々のものが用いられている。しかし，これまで工夫されたクラスター分析に用いられた類似性，非類似性の統計値の分布についてはいかなる仮定もおかず，したがって，それらの統計値の確率法則も利用せず個体の分類を考えるのが一般である。

Distribution-freeの考え方を利用する事は，その方法の応用の範囲を広げるが，統計的判断において，恣意的基準を用いざるを得ない弱点が存在する。勿論，恣意的判断基準に従って考察を進め得る事は観察の自由度を増大させ，研究の道具として都合のよい場合もある。

しかし，ある個体のあるクラスターへの所属決定の判断に，統計的推測を利用できると判断がより客観的になるといえる。しかも，多くの場合，類似性，または非類似性をあらわす統計値は標本統計値である。この事情からもクラスター分析に用いられる統計値の確率分布がわかっていれば，それを利用して個体の所属の判断を下すのが客観的結果を生む事になるであろう。

以上のことから，類似性を相関係数であらわした場合については，別の論文にまとめた[11]。そこで，本論文では非類似性を用いた場合について，統計的推測を分析過程において考慮する方法を開発することを目的とした。

1　非類似性統計値の制限

非類似性は一般に距離であらわされることが多い。すなわち，ユークリッドの距離，重みつきユークリッド距離，標準ユークリッド距離，マハラノビス汎距離，等々が用いられるが，これらの確率分布ははっきりしていない。

そこで，距離の性質をもっており，かつ確率分布の明確にわかっている統計値としては，平均値の差の検定に用いられるt，zの値，

$$t=\frac{|\bar{x}_1-\bar{x}_2|}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \tag{1}$$

$$z=\frac{|\bar{x}_1-\bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \tag{2}$$

などである。(1)は自由度(n_1+n_2-2)のt-分布(2)は正規分布N（0,1）に従うことがわかっている。したがって，個体間の距離を(1)，(2)の様な形で表わしうる場合，序論で述べた確率分布を利用してクラスター分類が可能になるであろう。

2　所属判定基準の理論的根拠

いま，n個の個体をx_1，x_2，x_3，……x_nとする。この各個体相互の非類似性（距離）を，標本数Nの資料を用いて，(1)または(2)式より求め，行列に整理したものを，

$$T=\begin{pmatrix} 0 & t_{12} & t_{13} & \cdots\cdots & t_{1n} \\ t_{21} & 0 & t_{23} & \cdots\cdots & t_{2n} \\ t_{31} & t_{32} & 0 & \cdots\cdots & t_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ t_{n1} & t_{n2} & t_{n3} & \cdots\cdots & 0 \end{pmatrix} \tag{3}$$

とする。ここで，標本数Nは適当に大であると仮定する。実際にはN≧40で十分である。この仮定は，この手法をプログラムにする時大変便利になる。

今，n個の個体の一部がl個のクラスターに，つぎの様に所属が決定し，l番目クラスターの所属個体を探索の途中とする。すなわち，

1番目クラスター；$(x_1^1,\ x_2^1,\ x_3^1,\ \cdots\cdots,\ x_{n_1}^1)$，
2番目クラスター；$(x_1^2,\ x_2^2,\ x_3^2,\ \cdots\cdots,\ x_{n2}^2)$，
3番目クラスター；$(x_1^3,\ x_2^3,\ x_3^3,\ \cdots\cdots,\ x_{n3}^3)$，
$\vdots\quad\vdots\quad\vdots\qquad\vdots$
l番目クラスター；$(x_1^l,\ x_2^l,\ x_3^l,\ \cdots\cdots,\ x_{nl}^l)$，

とlクラスターに，$\sum_{i=1}^{nl} n_i$個の個体の所属は決定し，$n-\sum_{i=1}^{nl} n_i=n_0$個の個体の所属が決定しておらず，$l$番目クラスターへの個体所属を検討する過程を考える。

このn_0の所属未決定の個体を

x'_1，x'_2，x'_3，……，x'_{n0}

とする。$(x'_j;\ j=1,\ 2,\ 3,\ \cdots\cdots,\ n_0)$の任意の個体$x'_k$と$l$番目クラスター所属の$n_l$この個体との距離を

$$t^l_{1k},\ t^l_{2k},\ t^l_{3k},\ \cdots\cdots,\ t^l_{nlk} \tag{4}$$

とする。これらt^l_{jk}；$j=1,\ 2,\ 3\cdots\cdots n_l$，$k=1,\ 2,\ 3,\ \cdots\cdots n_0$の$n_l n_0$このtの値は，標本数が適当に大である時には，正規分布；N（0,1）に近似的に従うことはわかっている[10]。

t^l_{jk}の平均値を$\bar{t}^l_{kn1}$とすれば，

$$\bar{t}^l_k=\frac{1}{n_l}\sum_{j=1}^{n_l} t^l_{jk}\ ;\ k=1,\ 2,\ 3,\ \cdots\cdots n_0 \tag{5}$$

である。ついで，$\bar{t}^l_k$；$k=1,\ 2,\ 3,\ \cdots\cdots n_0$の最小値を$\bar{t}^l_m$とする。すなわち，

$$\bar{t}^l_m=\min\ (\bar{t}^l_k\ ;\ k=1,\ 2,\ 3,\ \cdots\cdots n_0) \tag{6}$$

である。t^l_{jk}がN（0，1）に従うことから，$\bar{t}^l_m$はN$(0,\ \frac{1}{n_l})$に従うことは容易に導かれる*1。

---

*1，$t^l_{jk}\to N(0,\ 1)$，$\frac{1}{n_l}t^l_{jk}\to N(0,\ \frac{1}{n_l^2})$，であるから，平均値は，$\Sigma\frac{1}{n_l}t^l_{jk}\to N(0,\ \frac{n_l}{n_l^2})=N(0,\ \frac{1}{n_l})$

したがって，

$$\bar{t}_m^{l}=0$$

を帰無仮説とする時の検定は，z_0を

$$z_0=\frac{|\bar{t}_m^{l}-0|}{\sqrt{\frac{1}{n_l}}}=\sqrt{n_l}\,|\bar{t}_m^{l}| \tag{7}$$

とする時，$z_0<1.96(\alpha=0.05)$，$z_0<2.58(\alpha=0.01)$であれば，$\bar{t}_m^{l}=0$と推測される。したがって，x_m'をl番目クラスターに所属すると判断することが出来る。もし，この検定で$z_0\geqq1.96$，又は2.58である場合は，再びつぎの検討を行う。

l番目クラスター所属の個体相互間の距離全部が統計的に0であるものばかりでクラスターを構成するという事は理想的であるが，実際にはx_j^{l}；$j=1,\ 2,\ 3,\ \cdots\cdots n_l$，相互間の距離は存在する。これらの距離を，

$$(t_{12}^{l},t_{13}^{l},t_{14}^{l},\cdots\cdots,\ t_{(p-1)p}^{l}) \tag{8}$$

とする。ただし，$p=\frac{n_l(n_l-1)}{2}$である。この(8)と，(4)で与えらた$(x_1^{l},\ x_2^{l},\ x_3^{l},\ \cdots\cdots,\ x_{n\,l}^{l})$と$x_m$（今検討中の個体）との距離$(t_{1m}^{l},\ t_{2m}^{l},\ t_{3m}^{l},\ \cdots\cdots,\ t_{n_lm}^{l})$の平均値$\bar{t}_m^{l}$と(8)の平均値$\bar{t}^{l}$との差の有意性を検定する。この場合，(7)の場合と同様，$\bar{t}^{l}$は正規分布$N\left(0,\ \frac{2}{n_l(n_l-1)}\right)$に従うことが容易に導かれる*2。

それ故，$(\bar{t}^{l}-\bar{t}_m^{l})$は

$$N\left(0,\ \frac{2}{n_l(n_l-1)}+\frac{1}{n_l}\right)$$

$$=N\left(0,\ \frac{n_l+1}{n_l(n_l-1)}\right) \tag{9}$$

に従うことになる。

それ故，

$$z_0=\frac{|\bar{t}^{l}-\bar{t}_m^{l}|}{\sqrt{\frac{n_l+1}{n_l(n_l-1)}}} \tag{10}$$

とすれば，

$$z_0<1.96\text{又は}2.58 \tag{11}$$

が成立つ時，$\bar{t}^{l}$と$\bar{t}_m^{l}$には差がないと推測される。この場合，x_m'はl番目クラスターに所属すると判断する。

したがって，

(1)，$\bar{t}_m^{l}$が有意でない。

(2)，$|\bar{t}^{l}-\bar{t}_m^{l}|$が有意でない。

の2条件のいずれかが満足される場合には，x_m'を検討中のクラスターの要素として，クラスターを再構成し，

$$(x_1^{l},\ x_2^{l},\ \cdots\cdots,\ x_{n\,l}^{l},\ x_m')$$

をもって，l番目クラスターとする。再び，$(x_1',\ x_2',\ x_3',\ \cdots\cdots,\ x'_{no})$から$x_m'$を除いた個体について，これまでの手続を繰返し，(1)，(2)の2条件の両方が否定されるまで，クラスター要素の探索を続ける。

以上の2つの統計値を手がかりとして，個体のクラスター所属の判断を繰返し，すべての個体相互間の距離について行えば，非階層的クラスター構成が見出せる。階層的クラスター構成に到達するために，クラスター相互間の距離（tの値）を定義しておこう。

今2つのクラスター及び各クラスターの要素を次の通りとする。

C_p；$x_{p1},\ x_{p2},\ \cdots\cdots,\ x_{p\,l_p}$，

C_q；$x_{q1},\ x_{q2},\ \cdots\cdots,\ x_{q\,l_q}$，

C_pとC_qとの距離（tの値）はC_pのすべての要素とC_qのすべての要素の距離の平均値とする。

すなわち，

$$t_{pq}=\frac{1}{l_p l_q}\sum_{i}^{l_p}\sum_{j}^{l_q}t_{ij} \tag{12}$$

これらのクラスター相互間の距離は，

$$T_l=\begin{pmatrix}0 & t_{12} & t_{13}\cdots\cdots & t_{1m}\\ t_{21} & 0 & t_{23}\cdots\cdots & t_{2m}\\ \vdots & \vdots & \vdots & \vdots\\ t_{m1} & t_{m2} & t_{m3}\cdots\cdots & 0\end{pmatrix} \tag{13}$$

で与えられる。この距離行列(13)を，前述の(3)と見なして，以上の手続を再度繰返せば，第2段階目のクラスター分析が成就される。

さて，以上の考え方にもとづいた，クラスター所属の個体の探索は，距離の最小の両個体から出発して，順次に，残った個体のうちで，クラスターとの距離の最小のものを加えてゆくという方法である。したがって，クラスター所属判断の順位を考慮すれば，非階層的クラスター分析の場合でも階層的クラスターリングに到達する事が出来る。

に従う。

＊2，$t_{jk}^{l}\rightarrow N(0,\ 1)$，

$$\frac{2}{n_l(n_l-1)}t_{jk}^{l}\rightarrow N\left(0,\ \left(\frac{2}{n_l(n_l-1)}\right)^2\right)$$

$$\sum_{\substack{j,k\\ j<k}}^{n_l}\frac{2}{n_l(n_l-1)}t_{jk}^{l}\rightarrow N\left(0,\ \frac{2}{n_l(n_l-1)}\right)$$

さらに，第2段階，第3段階，……のクラスター分析の段階と個体のクラスター要素としての判断の順位を考慮すれば，従来の階層的クラスター分析の結果と同様な結果に到達しうる。しかし，すでに述べたクラスター所属の判断基準を適用する限り，最終段階で1つのクラスターに融合するとは限らない。しかし，これは全くの技術的問題で，残った個体間の距離のすべてが有意である場合には，最終段階でも1つのクラスターに融合しない事を意味している。したがって，この最終段階で，クラスター所属判定に用いた基準をはずし，距離の小なる個体相互からクラスター構成を考えれば，最終的には1つのクラスターにすべての個体が融合する事になり，デンドログラム（樹木図）が描ける事になる。しかし，この事は，本方法にとって重要なことではない。何故なら，本方法は，従来のクラスター分析におけるクラスター数，クラスター要素の最終的決定の恣意性に対する批判から工夫されたものであるからである。

3　アルゴリズム

1)，n個の個体相互の距離行列を(1)，(2)などの統計値を利用して求める。

2)，この距離行列の要素の中で最小の値を見出し，これをt_{ij}とする時，$t_{ij}<1.96(\alpha=0.05)$又は$2.58(\alpha=0.01)$を検討する。これが満足される時，x_iとx_jはクラスターを構成すると判断する。これが満足されない時は，x_iとx_j及び他のすべての個体の各々が別々のクラスターを構成すると判断し，クラスターの個体分類を停止する。

3)，x_iとx_jがクラスターを構成すると判断された場合は，このクラスターに所属する他の要素を探索する。

x_i，x_jと，これら以外の要素との距離の平均値を，$l=1$，$n_l=2$とおいて，

$$\bar{t}_r^l=\frac{1}{n_l}\sum_{p=1}^{n_l} t_{pr},\quad r=1,\ 2,\ 3\cdots\cdots n,\ r\neq i,\ j$$

を求める。

4)，$\bar{t}_r$；r＝1，2，3，……n，r≠i，jの最小値を求める。

$\bar{t}_m^l=\min(\bar{t}_r$；r＝1，2，3，……n，r≠i，j)

5)，$z=\sqrt{n_l}\,|\,\bar{t}_m^l\,|<1.96(\alpha=0.05)$又は$2.58(\alpha=0.01)$であれば，$x_m$はクラスターに所属すると判断する。この判断が下された時は，$n_l=n_l+1$として3)のステップに戻る。

6)，5)が否定された時は，

$$z=\frac{\sqrt{\ln_l(n_l-1)}\,|\,\bar{t}^l-\bar{t}_m^i\,|}{\sqrt{\ln_l+1}}$$

$$\bar{t}^l=\frac{2}{n_l(n_l-1)}\sum_{\substack{p=1\\p\neq q}}^{n_l}\sum_{q=1}^{n_l} t_{pq}$$

；クラスター要素相互間の距離の平均値

を求め，$z<1.96(\alpha=0.05)$又は$2.58(\alpha=0.01)$であれば，x_mはクラスターに所属すると判断する。そして，$n_l=n_l+1$として，3)にもどる。

7)，6)で，$z<1.96(\alpha=0.05)$又は$2.58(\alpha=0.05)$が成立しない場合は，このクラスターに属する個体の探索を止める。

8)，残った個体の数が2に等しいか，または2より大である場合には，新しい別のクラスターの構成に移る。すなわち，クラスターへの所属が，未だ決定されていない個体相互間の距離の最小の個体を見出す。

9)，この距離について，$z_0=|\,t_{kl}\,|<1.96(\alpha=0.05)$，又は$2.58(\alpha=0.01)$を検討し，成立てばこれら2個体は1つのクラスターを構成すると判断する。成立たない時は，残った個体はそれぞれ別々のクラスターを構成すると判断する。

10)，ここから，3)にもどる。

11)，8)において，残った個体の数が2より小，つまり1である時には，この残った1つの個体が1つのクラスターを構成すると判断する。

12)，2)，9)，及び11)において，クラスタリングの手続が終ってループを抜け出た場合には，結果を出力する。

13)，階層的クラスター分析を実行する場合には，つぎに，前段階で得られたクラスターの相互間距離を前節の(12)式で求め，距離行列T_l(13)を求める。

14)，再び2)の所へもどって，計算を繰返す。

以上のアルゴリズムをフローチャートで示したものを附録(1)に示してある。

4　個体相互間の距離が与えられた時の個体の布置の推定

3のアルゴリズムに従って個体は相互の距離を手がかりに分類することは出来るが，分類された個体を視覚的に確める事が出来れば，クラスターの意味する所がさらに明確になるであろう。そのため，距離行列から，各個体の適当な次元の空間における座標を決定する事を工夫する。

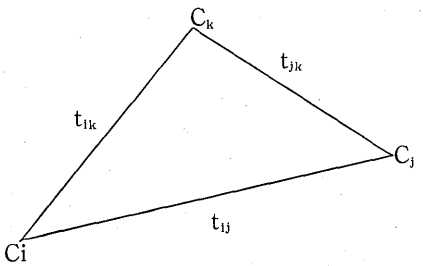
nこの個体のうち，任意の3個体をC_i，C_j，C_k

とし，m次元直交空間における座標をそれぞれ，

C_i；$(a_{i1}, a_{i2}, a_{i3}, \cdots\cdots, a_{im})$

C_j；$(a_{j1}, a_{j2}, a_{j3}, \cdots\cdots, a_{jm})$

C_k；$(a_{k1}, a_{k2}, a_{k3}, \cdots\cdots, a_{km})$とし，



これらの3個体相互間の距離をt_{ij}，t_{ik}，t_{jk}とする。これらの距離は既知である。t_{ij}，t_{ik}，t_{jk}をユークリッドの距離と仮定すれば，

$$t_{ij}^2=\sum_{p=1}^{m}(a_{ip}-a_{jp})^2 \tag{14}$$

$$t_{ik}^2=\sum_{p=1}^{m}(a_{ip}-a_{kp})^2 \tag{15}$$

$$t_{jk}^2=\sum_{p=1}^{m}(a_{jp}-a_{kp})^2 \tag{16}$$

とあらわされる。ここで，C_iに原点を移動すれば，C_iの座標は

C_i；(0，0，0，……0)

となる。かつ，

$$t_{ij}^2=\sum_{p=1}^{m}a_{jp}^2,\quad t_{ik}^2=\sum_{p=1}^{m}a_{kp}^2$$

$$t_{jk}^2=\sum_{p=1}^{m}(a_{jp}-a_{kp})^2$$

と簡単になる。したがって，

$$t_{jk}^2=\sum_{p=1}^{m}a_{jp}^2+\sum_{p=1}^{1}a_{kp}^2-2\sum_{p=1}^{m}a_{jp}a_{kp}$$

したがって，

$$\sum_{p=1}^{m}a_{jp}a_{kp}=\frac{1}{2}(t_{ij}^2+t_{ik}^2-t_{jk}^2)$$

となり，m次元直交空間におけるC_jとC_kの座標の積和，すなわち，ベクトルC_jとC_kとの内積は$\frac{1}{2}(t_{ij}^2+t_{ik}^2-t_{jk}^2)$で与えられることがわかる。したがって，

$$v_{jk}=\frac{1}{2}(t_{ij}^2-t_{ik}^2-t_{jk}^2) \tag{17}$$

j，k＝1，2，3，……，n，j，k≠i

とする時，

$$\boldsymbol{V}=\begin{pmatrix} v_{11} & v_{12} & v_{13} & \cdots\cdots & v_{1n} \\ v_{21} & v_{22} & v_{23} & \cdots\cdots & v_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ v_{n1} & v_{n2} & v_{n3} & \cdots\cdots & v_{nn} \end{pmatrix}=(v_{jk},\ j,\ k\neq i)$$

は(n－1)次の対称行列であり，かつ，v_{jk}はC_jとC_kの内積であるから，$\boldsymbol{V}$は(n－1)このベクトル相互の内積行列である。

したがって，a_{jp}；j＝1，2，3……n，p＝1，2，3……mを要素とする行列を

$$\boldsymbol{A}=\begin{pmatrix} a_{11} & a_{12} & \cdots\cdots & a_{1m} \\ a_{21} & a_{22} & \cdots\cdots & a_{2m} \\ a_{31} & a_{32} & \cdots\cdots & a_{3m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots\cdots & a_{nm} \end{pmatrix}=(a_{jk},\ j\neq i)$$

とする時，$\boldsymbol{AA}'$の要素は

$$\sum_{p=1}^{m}a_{jp}a_{kp}=\begin{cases}\sum_{p=1}^{m}a_{jp}^2\ ;\ j=k\ ;\ j\neq i \\ \sum_{p=1}^{m}a_{jp}a_{kp}\ ;\ j\neq k\ ;\ j,\ k\neq i\end{cases} \tag{18}$$

である。したがって，$\boldsymbol{AA}'$はV行列に等しい。したがって，$\boldsymbol{V}$の固有値を大きさの順に列べて，

$$\lambda_l\geqq\lambda_2\geqq\lambda_3\cdots\cdots\lambda_m \tag{19}$$

$$m\leqq(n-1)$$

とし，これらを対角線要素とする行列を

$$\boldsymbol{E}=\begin{pmatrix} \lambda_l & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \ddots\ \lambda_m \end{pmatrix} \tag{20}$$

とし，λj；j＝1，2，3，……mに対応する固有ベクターを列要素とする行列を

$$\boldsymbol{B}=\begin{pmatrix} b_{11} & b_{12} & b_{13} & \cdots\cdots & b_{1m} \\ b_{21} & b_{22} & b_{23} & \cdots\cdots & b_{2m} \\ b_{31} & b_{32} & b_{33} & \cdots\cdots & b_{3m} \\ \vdots & \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots\cdots & b_{nm} \end{pmatrix}=(b_{jk},\ j\neq i) \tag{21}$$

とする時，

$$\boldsymbol{A}=\boldsymbol{B}\boldsymbol{E}^{\frac{1}{2}}\ ^{*3} \tag{22}$$

で与えられる。

したがって，C_lを原点として，残りの$(n-1)$この個体のm次元空間における座標が決定される。この場合，原点のとり方はn通りであり，したがって，n通りの解が得られる事になる。しかし，これらの解相互は座標系の平行移動，回転によって変換可能である事から，各個体の相対的布置はかわらないが，C_lを原点とした場合の解をA_lとすれば，A_l；i＝1，2，3，……nの平均をもって解とする方法がよく用いられる[7]。また，原点を最初から各個体の重心に固定して，内積行列$\boldsymbol{V}$を求め，この(j，k)要素を，

$$v_{jk}=\frac{1}{2}\{(\frac{1}{n}\Sigma_j t_{jk}^2+\frac{1}{n}\Sigma_k t_{jk}^2)-(\frac{1}{n^2}\Sigma_j\Sigma_k t_{jk}^2+t_{jk}^2)\} \quad (23)$$

であらわして，解けば，一義的解が得られる[8]。

また，(17)式において，iは1，2，3，……nを取る事が出来る事から

$$v_{jk}=\frac{1}{2}(t_{ij}^2+t_{ik}^2-t_{jk}^2),\ i=1,\ 2,\ 3,\ \cdots\cdots n$$

であり，

$$nv_{jk}=\frac{1}{2}\ (\sum_{i=1}^{n} t_{ij}^2+\sum_{i=1}^{n} t_{ik}-nt_{jk}^2)$$

である。したがって，

$$v_{jk}=\frac{1}{2}(\frac{1}{n}\sum_{i=1}^{n} t_{ij}^2+\frac{1}{n}\sum_{i=1}^{n} t_{ik}^2-t_{jk}^2) \quad (24)$$

が導かれる。A_lの平均をもって解とする方法は，n回の固有値，固有ベクターの計算を含む事から膨大な計算が必要となるが，(24)式を用いる場合は一回の固有値，固有ベクターの計算ですむ。

(17)，(23)はYoung，GとHousholder，A.S.(1938)[9]によって提案されたものであり，(24)は筆者によるものである。

4　運動部の分類への適用

12の大学男子運動部；1．陸上競技（走，跳），2．バスケット・ボール，3．バトミントン，4．テニス，5．剣道，6．サッカー，7．水泳，8．ハンドボール，9．弓道，10．野球，11．ラグビー，12．バレー・ボール，の部員に1．身長，2．体重，3．背筋力，4．懸垂，5．垂直跳，6．肺活量，7．5分間走，8．最大酸素摂取量，9．体前屈，10．全身反応時間，11．100m走，12．ソフトボール投の12項目の測定を行い，各運動部の標本数が40を越えるように資料を収集した。各変量毎に，全標本をプールした時の平均値，標準偏差を用いて，測定値を標準化し，この標準化された観測値を分析のための資料とした。

i番目運動部のk番目変量の平均値をx_{ik}とする。iとj運動部の距離をつぎのように求めた。

i運動部；x_{i1}，$x_{i2}\cdots\cdots x_{im}$

j運動部；x_{j1}，$x_{j2}\cdots\cdots x_{jm}$

$$d_{ijk}=x_{ik}-x_{jk}\ ;\ k=1,\ 2,\ 3,\ \cdots\cdots m$$

$$\bar{d}_{ij}=\frac{1}{m}\sum_{k=1}^{m} d_{ijk}$$

$$S_{ij}^2=\frac{1}{m-1}\Sigma(d_{ijk}-\bar{d}_{ij})^2$$

$$t_{ij}=\frac{|\bar{d}_{ij}|}{\sqrt{\frac{S_{ij}^2}{m}}}=\frac{\sqrt{m}\,|\bar{d}_{ij}|}{S_{ij}} \quad (25)$$

t_{ij}はx_{ik}，x_{jk}がそれぞれ独立で正規分布に従うと仮定できる時，自由度$(m-1)$のt分布に従う。求められたt_{ik}はx_{ik}，x_{jk}が異なる運動部の観測値であり，かつ変量毎に標準化されている事から，上の仮定に従うと考えてよい。

そこで，上式(25)を用いて，得られた距離行列はTable 1の通りである。この行列の要素は各運動部間での有意差を示すtの値である。

この距離行列に本手法を適用した結果が，Table 2に示されている。Team#の欄の番号の順序は各クラスターに所属すると判断された順序と一致している。したがって，cluster 1についてみると，まず，6．サッカー，8．ハンドボールのチームが類似度が高くクラスターを構成し，そのクラスターに，バスケットボールが所属すると判断され，ついで，バレーボールが所属すると判断された事を示す。そして，クラスター2は陸上競技と野球，クラスター3は剣道と弓道，クラスター4はバドミントンとテニス，クラスター5は水泳とラグビーと，以上，12運動部は用いた体力変量からは5つのクラスターに分類されると推測された。

ついで，この5つのクラスター相互間の距離を(12)式を用いて計算したものがTable 3である。これに，再び本法を適用した結果が，Table 4の通り

＊3，$\boldsymbol{C}=\boldsymbol{BEB}'=(\boldsymbol{BE}^{\frac{1}{2}}(\boldsymbol{BE}^{\frac{1}{2}})'=\boldsymbol{AA}'$

Table 1 Distance matrix between 12 college sports teams

team#											
1											
2	2.64										
3	3.46	0.83									
4	4.51	2.06	1.40								
5	2.39	1.18	1.71	2.91							
6	2.71	0.47	0.90	1.90	1.23						
7	0.87	3.42	4.24	5.29	2.96	3.47					
8	2.87	0.51	0.78	1.96	0.99	0.44	3.59				
9	2.01	1.81	2.41	3.36	1.04	1.72	2.43	1.68			
10	0.66	2.44	3.24	4.20	2.12	2.43	1.15	2.59	1.53		
11	0.76	1.97	2.79	3.78	1.94	2.01	1.59	2.23	1.67	0.72	
12	2.83	1.01	1.28	2.33	0.80	0.83	3.47	0.62	1.49	2.53	2.25

Note; team # : 1. Track & field, 2. Basketball, 3. Badmington, 4. Tennis,
5. Kendo, 6. Soccer, 7. Swimming, 8. Handball,
9. Kyudo, 10. Baseball, 11. Rugby, 12. Volleyball.

Table 2 Clusters of 12 teams at the 1st stage of clustering

Cluster	No. of teams	Team #			
Cluster 1	4	6	8	2	12
Cluster 2	2	1	10		
Cluster 3	2	5	9		
Cluster 4	2	3	4		
Cluster 5	2	7	11		

Note; Team # corresponds to the number of team # shown in table 1.

Table 3 Distance matrix among clusters identified at the first stage of clustering

Cluster#				
1				
2	2.63			
3	1.36	2.01		
4	1.51	3.85	2.60	
5	2.80	0.87	2.525	4.02

Note; Cluster# coresponds to the cluster produced at the 1st stage of clustering

Table 4 Clusters identified at the 2nd stage of clustering

Cluster#	No. of clusters	Cluster# *	
1	2	2	5
2	2	1	3
3	1	4	

Note; Cluster # * corresponds to the cluster # produced at the 1st stage of clustering

Table 5 Distance matrix among the clusters produced at the 2nd stage of clustering

Cluster#		
1		
2	2.42	
3	3.94	2.05

Note; Cluster# corresponds to the cluster# identified at the 2nd stage of clustering

Table 6 Clusters identified at the 3rd stage of clustering

Cluster#	No. of cluster	Cluster# *
1	1	2
2	1	3
3	1	1

Note; Cluster # * corresponds to the cluster # produced at the 2nd stage of clustering

で，3つのクラスターに分類され，各クラスターに属する第1段クラスター分析で得られたクラスターの番号はcluster#＊の欄の通りである。

さらに，この3クラスター相互間の距離をTable.3を用いて，(12)式で求めたものが，Table 5である。Table 5の各tの値は1.96より大である事から，これらの3クラスター相互は有意な差があると推定される。本手法を適用すると，Table 6の通りであり，第2段階で得られた3クラスターは，本手法ではこれ以上融合する事はないと判断された。

強制的に融合させるためには，本手法における統計的有意性(アルゴリズム；5)，差の有意性(アルゴリズム；6）の検定を除けば，Table 6のcluster#＊の欄に示されている番号2，3，1の順序で融合する事になる。

以上，本手法を最初の1回の適用で計算を打ち切れば非階層的クラスター構成が得られ，つぎの通りであった。

cluster 1；サッカー，ハンドボール，バスケットボール，バレーボール
cluster 2；陸上競技，野球
cluster 3；剣道，弓道
cluster 4；バドミントン，テニス
cluster 5；水泳，ラグビー

ついで，3回のクラスター分析の結果と，クラスター所属判断の順序を考慮して，階層的クラスター構成を示すデンドログラムを描くとFig.1の通りである。

つぎに，これら12運動部の適当な次元における布置を求め，クラスター分析の結果を視覚的に考察することにする。

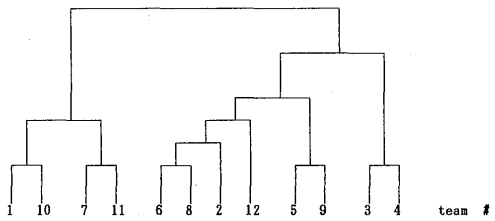


Fig.1 Dendrogram; hierarchical clustering of 12 university sports teams

Team#; 1. Track & field, 2. Basketball, 3. Badmington, 4. Tennis, 5. Kendo, 6. Soccer, 7. Swimming, 8. Handball, 9. Kyudo, 10. Baseball, 11. Rugby, 12. Volleyball,

同一の距離行列（Table 1）から，(17)，(24)，(23)式を用いて求められた内積行列はTable 7，Table 8，Table 9である。これらの正方対称行列の固有値，固有ベクターを求め，(22)式から，Aすなわち，座標行列を求めた。Table 10が示すように2軸(A_1，A_2)でトレースの88%以上が，いずれの方法で求められた内積行列の場合も説明される。したがって，2次元空間で十分これらの運動部の布置は説明されると考えられる。なお，これらの2軸は，両軸のトレースに対する貢献量が等しくなるように回転してある。この2軸に対する各運動部の座標はTable 11に示されている。これらの座標を資料として各運動部の布置状況をグラフに示し

Table 7 Inner-produc tmatrix estimated with procedure I.

Individual											
1	7.39	.31	−1.91	−4.68	1.58	9.35	−.35	3.14	6.44	5.44	.01
2	.31	.22	.17	−.21	.16	.29	.07	−.05	.08	.19	−.06
3	−1.91	.17	.82	1.24	−.3	−2.55	.2	−1.03	−1.9	−1.46	−.07
4	−4.68	−.21	1.24	3.63	−1.68	−6.18	−.01	−2.35	−4.05	−3.3	−.57
5	1.58	.16	−.3	−1.68	1.52	2.42	.36	1.7	1.47	.9	.79
6	9.35	.29	−2.55	−6.18	2.42	12.08	−.32	4.57	8.34	6.8	.36
7	−.35	.07	.2	−.01	.36	−.32	.19	.16	−.31	−.36	.25
8	3.14	−.05	−1.03	−2.35	−1.7	4.57	.16	2.96	3.27	2.12	.72
9	6.44	.08	−1.9	−4.05	1.47	8.34	−.31	3.27	5.94	4.74	.1
10	5.44	.19	−1.46	−3.3	.9	6.8	−.36	2.12	4.74	4.07	−.16
11	.01	−.06	−.07	−.57	.79	.36	.25	.72	.1	−.16	.69

$v_{jk}=\frac{1}{2}\left(t_{ij}^2+t_{ik}^2-t_{jk}^2\right)$, i=6 fixed

Table 8 Inner-product matrix estimated with procedure II.

Individual												
1	6.22	1.28	−.28	−2.11	1.9	1.06	7.48	.73	2.94	5.5	4.92	.9
2	1.28	3.31	3.93	4.5	2.61	3.19	.55	3.27	1.87	1.26	1.8	2.96
3	−.28	3.93	5.24	6.61	2.81	3.86	−1.63	4.06	1.56	−.05	.81	3.61
4	−2.11	4.5	6.61	9.95	2.38	4.8	−4.3	4.8	1.18	−1.26	−.08	4.06
5	1.9	2.61	2.81	2.38	3.32	2.55	2.03	2.91	2.98	2.01	1.87	3.16
6	1.06	3.19	3.86	4.8	2.55	3.29	.36	3.3	2.02	1.28	1.71	3.12
7	7.48	.55	−1.63	−4.3	2.03	.36	9.51	.05	3.66	6.69	5.58	.55
8	.73	3.27	4.06	4.8	2.91	3.3	.05	3.5	2.2	.97	1.36	3.38
9	2.94	1.87	1.56	1.18	2.98	2.02	3.66	2.2	3.72	3.28	2.56	2.57
10	5.5	1.26	−.05	−1.26	2.01	1.28	6.69	.97	3.28	5.21	4.44	1.21
11	4.92	1.8	.81	−.08	1.87	1.71	5.58	1.36	2.56	4.44	4.2	1.37
12	.9	2.96	3.61	4.06	3.16	3.12	.55	3.38	2.57	1.21	1.37	3.64

$$v_{jk}=\frac{1}{2}\left(\frac{1}{n}\sum_i t_{ij}^2+\frac{1}{n}\sum_i t_{ik}^2-t_{jk}^2\right)$$

Table 9 Inner-product matrix estimated with procedure III.

Individual												
1	6.07	1.12	−.44	−2.26	1.75	.91	7.33	.57	2.79	5.34	4.77	.75
2	1.12	3.16	3.77	4.34	2.46	3.04	.4	3.12	1.72	1.11	1.65	2.81
3	−.44	3.77	5.09	6.46	2.66	3.7	−1.78	3.91	1.4	−.2	.66	3.46
4	−2.26	4.34	6.46	9.8	2.23	4.65	−4.46	4.65	1.03	−1.41	−.23	3.91
5	1.75	2.46	2.66	2.23	3.17	2.39	1.88	2.76	2.83	1.85	1.71	3.01
6	.91	3.04	3.7	4.65	2.39	3.14	.21	3.14	1.87	1.13	1.55	2.97
7	7.33	.4	−1.78	−4.46	1.88	.21	9.36	−.11	3.51	6.54	5.43	.4
8	.57	3.12	3.91	4.65	2.76	3.14	−.11	3.35	2.04	.82	1.2	3.23
9	2.79	1.72	1.4	1.03	2.83	1.87	3.51	2.04	3.57	3.13	2.4	2.42
10	5.34	1.11	−.2	−1.41	1.85	1.13	6.54	.82	3.13	5.05	4.29	1.05
11	4.77	1.65	.66	−.23	1.71	1.55	5.43	1.2	2.4	4.29	4.05	1.22
12	.75	2.81	3.46	3.91	3.01	2.97	.4	3.23	2.42	1.05	1.22	3.49

$$v_{jk}=\frac{1}{2}\left[\left(\frac{1}{n}\sum_j t_{jk}^2+\frac{1}{n}\sum_k t_{jk}^2\right)-\left(\frac{1}{n}\sum_j\sum_k t_{jk}^2+t_{jk}^2\right)\right]$$

Table 10 The characteristics of two axes determined by three different procedures to evaluate the inner-product matrix.

Procedure	Trace of inner-product matrix	Amount of cont. of two axes A1	A2	Total amount of contribution	Total degree of contribution
I	39.568	18.967	18.883	37.850	95.66 %
II	30.587	16.264	13.342	29.606	96.79 %
III	32.347	14.981	13.497	28.478	88.04 %

Note; The (j,k) element of the 9 inner-product matrix was estimated with the following formulae;

$v_{jk}=\frac{1}{2}\left(t_{ij}^2+t_{ik}^2-t_{jk}^2\right)$, with i=6 fixed in Procedure I,

$v_{jk}=\frac{1}{2}\left(\frac{1}{n}\sum t_{ij}^2+\frac{1}{n}\sum t_{ik}^2-t_{jk}^2\right)$ in Procedure II, and

$v_{jk}=\frac{1}{2}\left[\left(\frac{1}{n}\sum_j t_{jk}^2+\frac{1}{n}\sum_k t_{jk}^2\right)-\left(\frac{1}{n^2}\sum_j\sum_k t_{jk}^2+t_{jk}^2\right)\right]$ in Procedure III.

Table 11 Coordinates of teams in two dimensional space constructed from the inner-product matrix

	Procedure I		Procedure II		Procedure III	
Team #	A1	A2	A1	A2	A1	A2
1	2.166	1.626	1.641	0.972	1.431	1.705
2	0.066	0.042	− 0.406	− 0.685	− 0.289	− 0.283
3	− 0.591	− 0.461	− 1.050	− 1.207	− 0.749	− 0.949
4	− 1.011	− 1.504	− 1.449	− 2.248	− 1.769	− 1.378
5	− 0.188	1.183	− 0.914	0.259	0.352	− 0.127
6	0.0	0.0	− 0.489	− 0.690	− 0.375	− 0.287
7	2.486	2.424	1.942	1.785	2.734	1.146
8	− 0.329	0.210	− 0.820	− 0.509	− 0.343	− 0.493
9	0.299	1.581	− 0.219	0.876	0.658	0.459
10	1.806	1.594	1.485	0.724	0.296	2.018
11	1.737	1.019	1.225	0.367	0.878	1.248
12	− 0.456	0.607	− 0.965	− 0.074	− 0.107	− 0.461

Note; 1) A1 and A2 stand for axis 1 and 2.

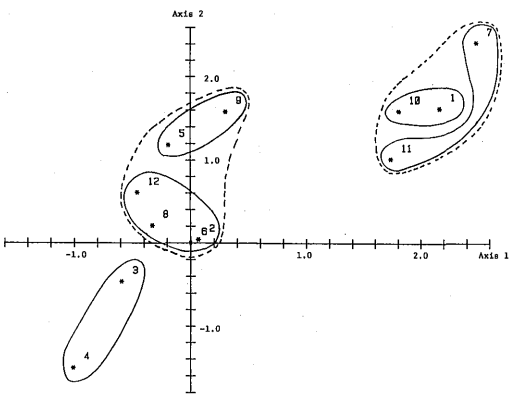


Fig.2 Team configuration in 2 dimensional space constructed from the inner-product matrix estimated with distance matrix; boy Procedure I

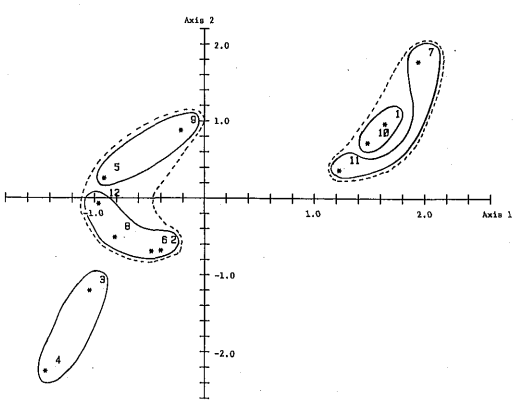


Fig.3 Team configuration in 2 dimensional space constructed from the inner-product matrix estimated with distance matrix; boy Procedure II

たものがFig.2, Fig.3, Fig.4である。Fig.1のデンドログラムとFig.2～Fig.4を考えあわせ，第2段階のクラスター分析で得られた3クラスターを解とするのが適当と考えられた。すなわち，

第1クラスター；陸上競技部，野球部，水泳部，ラグビー部

第2クラスター；バスケットボール部，サッカー部，ハンドボール部，バレーボール部，剣道部，弓道部

第3クラスター；バドミントン部，テニス部

とりあげた12運動部は12項目の体力変量からみる場合上記の3クラスターに分類されると推測された。

5 まとめと討論

平均値の差異の検定に利用される諸統計量をもって，個体間の距離を評価し，その統計量の確率分布を利用して非階層的クラスター分析の手法を工夫した。従来の多くのクラスター分析手法では，個体間の距離のみを手がかりとして，資料及び求められた距離について正規性，線型性などの統計的仮定を全く必要としない[5]。この点は従来の諸方法の大きな長所である。しかし，個体のク

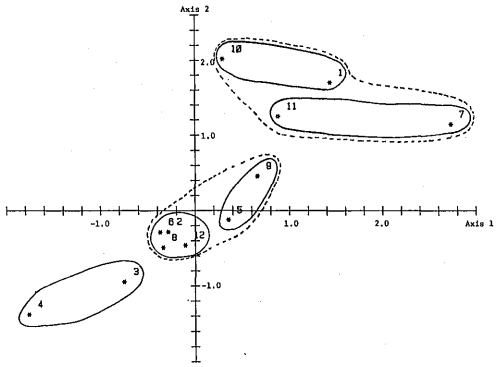


Fig.4 Team configuration in 2 dimensional space constructed from the inner-product matrix estimated with distance matrix; boy Procedure III

ラスター所属の判断は恣意的，直観的になされねばならず，この点は大きな短所であろう。この点を，本手法は距離をあらわす統計量の分布を利用して解決した。しかし，それだけ，距離統計量の分布が既知である必要があり，適用の制限が従来の手法より大であるといわなければならない。本手法では，クラスター内個体の分散についてのminimum variance条件の検討が含まれていない。これはクラスターの妥当性の1つの基準であるが[6]，本手法は，クラスター内の個体間の距離の差の有意性の有無を導入した。

本手法では，クラスター間の距離を(12)式で定義し，前段階で得られたクラスター相互間の距離を求め，クラスター相互間距離行列を作り，これに本手法を再度適用して，クラスターを個体とみなしてクラスター分析を行い，さらに，個体のクラスター所属判断の順序を考慮すれば，従来の階層的クラスター分析を行うことが出来る。したがって，本手法は，本来非階層的クラスターの解をねらって工夫されたが，非階層的解を繰返すことによって，階層的解に到達することも可能である。従来の非階層的方法では，クラスターに関する適当な分割を初期条件として与えて行うことが多い(再配置法，丘登り法，強制移動法等)[6]。しかし，本手法はこの必要はない。

ついで，得られたクラスター分類を可視的に考察するため，与えられた距離行列から，各個体の適当な空間における布置を決定する手法を工夫した。結局，Young and Housholderの考えに到達したが，2つの個体の内積行列の作り方について，(17)[7]，(24)式とよく用いられている(23)[8]式の3式が使用可能であることを示した。

以上の理論的根拠にもとづいて，12体力変量に関する12運動部集団の標準化された資料を用いて12運動部のクラスター分類を行い，かつ，2次元空間における各運動部の布置を決定した。クラスター分析より得られたデンドログラムと各運動部の布置から，クラスターは，第2段階クラスター分析によって得られた3クラスターに分類されるとするのが適当と推測された。

デンドログラム作成まではマイクロ・コンピューターでもそれほど時間はかからないが，布置の決定には，固有値，固有ベクターの計算，軸の回転の計算が含まれるのでマイクロ・コンピューターではかなりの時間がかかる。また，軸の回転には，Normal varimiax基準等の因子分析によく用いられる回転法のアルゴリズムを応用するのも一つの方法である。とくに，軸を解釈したい場合は有効である。しかし，軸の解釈を必要としない場合には，すべての軸の貢献量が等しくなる様に回転する事がよいであろう。これは，個体の布置を示す座標の各軸の尺度が等しくなるからである。この場合，i, j, 2軸の回転角は次式で与えられる。

$$\frac{(\sum_k a_{kj}^2-\sum_k a_{ki}^2)}{2\sum_k a_{ki}a_{kj}}=\tan 2\theta_{ij}$$

$$\theta_{ij}=\frac{1}{2}\tan^{-1}\left(\frac{(\sum_k a_{kj}^2-\sum_k a_{ki}^2)}{2\sum_k a_{ki}a_{kj}}\right) *4$$

必要な諸計算は，デンドログラムまでは，HP-85マイクロコンピューターで，布置の決定は本学学術情報センター，FACOM-M380で行なわれた。

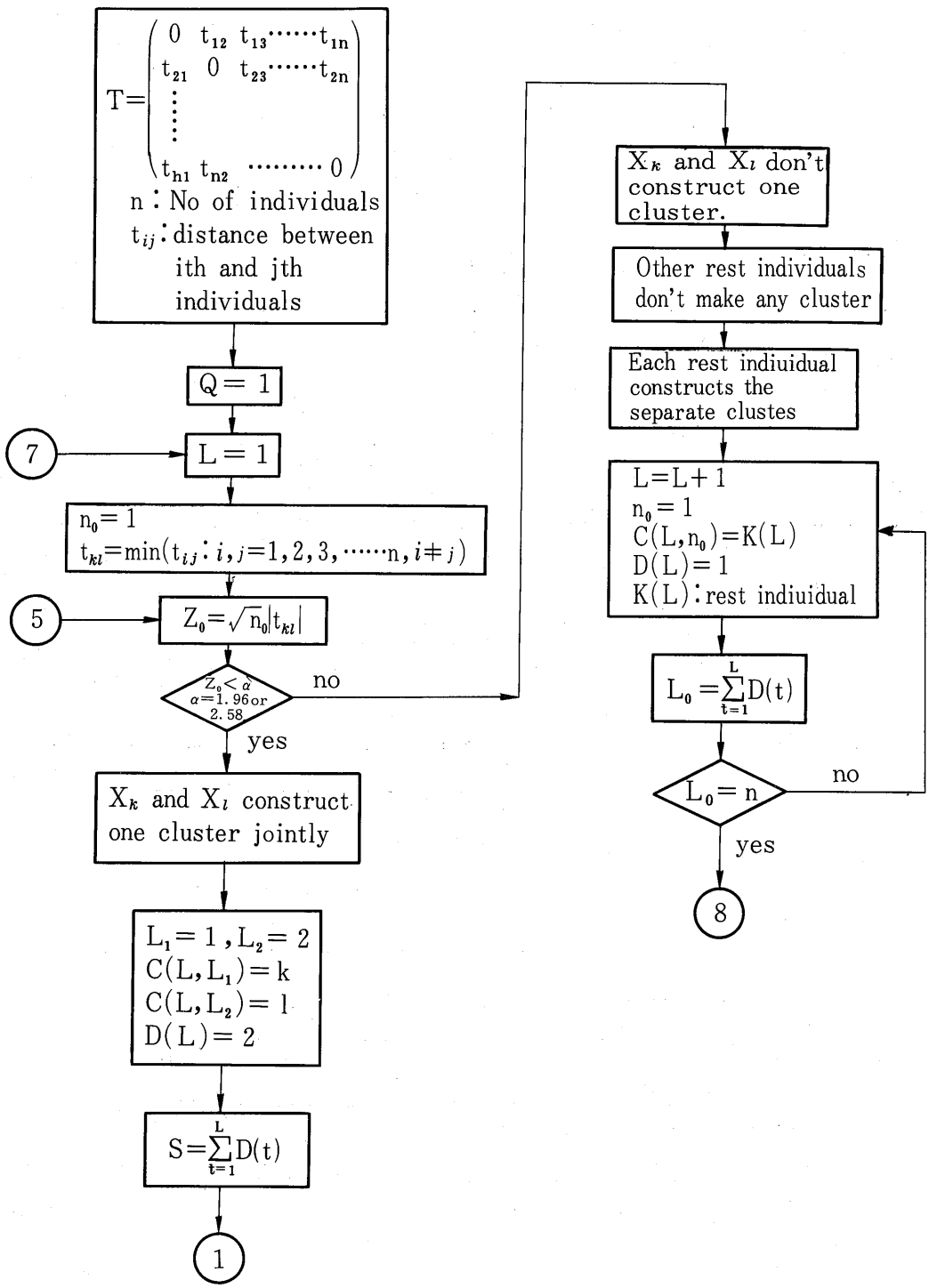
＊4，導出は附録(2)参照

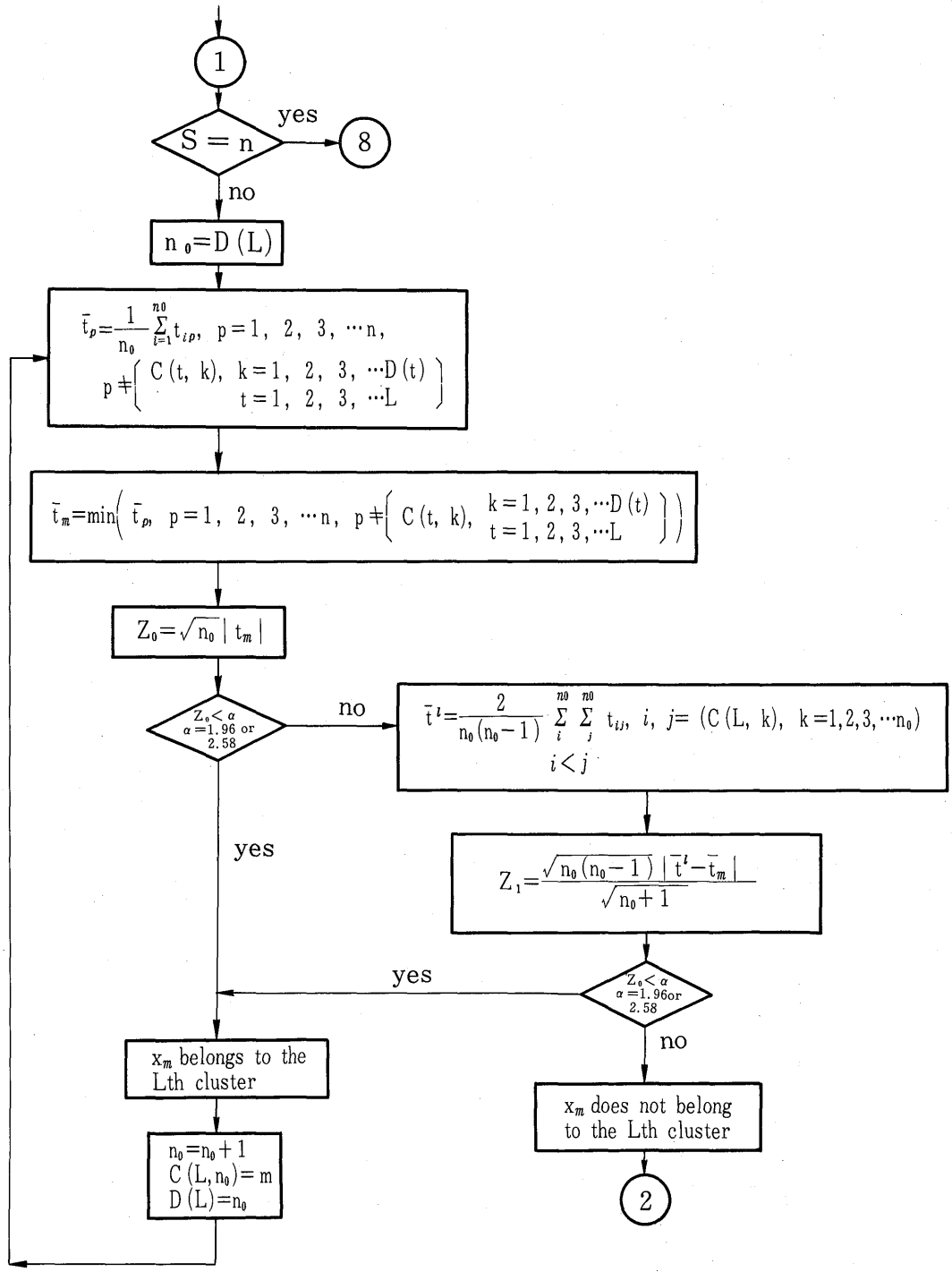
引用・参考文献

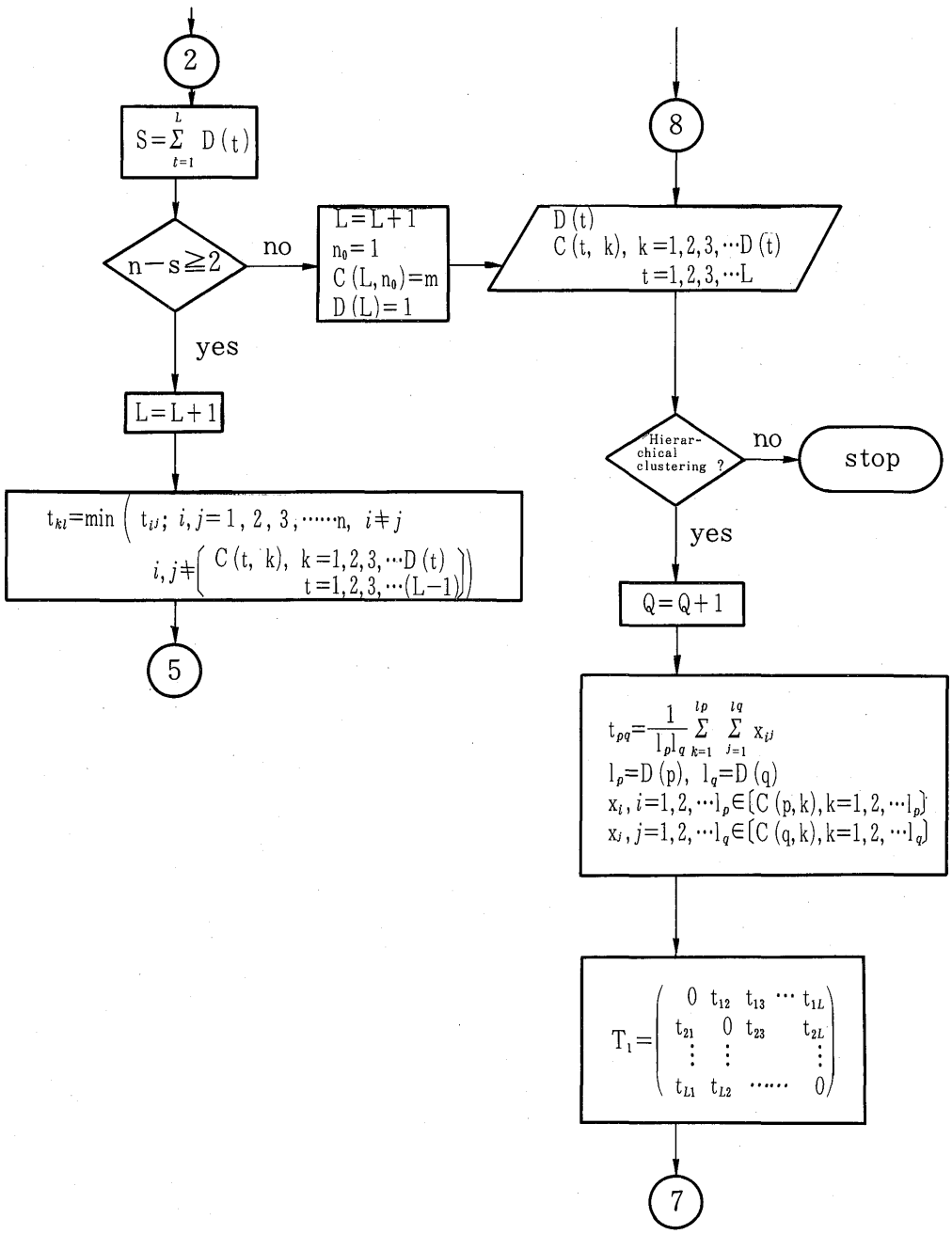
1) Fiedman, H. P. and Rubin, J., On some invariant criteria for grouping data, J. of American statistical association, 62, 1159-1178, 1967.
2) Gower, J. C., A comparison of some methods of cluster analysis, Biometrics, 23, 623-637, 1967.
3) Harman, H.,Modern factor analysis, The Univ. of Chicago press, 128-131, 1962.
4) 松浦義行，行動科学における因子分析法，不昧堂，

124-128, 1972.
5) 奥野忠一, 久米　均, 芳賀敏郎, 吉沢　正, 多変量解析法, 日科技連, 393, 1972.
6) 前掲書, 398-400.
7) 吉田正昭, 心理統計学, 丸善, 221, 1976.
8) 前掲書, 222.
9) Young, G. and Housholder, A. S., Discussions of a set of points in terms of their mutual distances, Psychometrika, 3, 19-22, 1938.
10) Wilks, S. S., Mathematical statistics, John Wiley Sons, 189, 1962.
11) 松浦義行, 類似度（相関係数）の統計的有意性を用いたクラスター分析手法, 未発表.

附録(1)　クラスター分析フローチャート




1
S = n
yes
8
no
n 0=D (L)
$\bar{t}_p=\frac{1}{n_0}\sum_{i=1}^{n0} t_{ip}$, p=1, 2, 3, ⋯n,
$p\neq\left\{\begin{array}{ll} C(t,k), & k=1,2,3,\cdots D(t) \\ & t=1,2,3,\cdots L \end{array}\right\}$
$\bar{t}_m=\min\left(\bar{t}_p,\ p=1,2,3,\cdots n,\ p\neq\left\{C(t,k),\ \begin{array}{l} k=1,2,3,\cdots D(t) \\ t=1,2,3,\cdots L \end{array}\right\}\right)$
$Z_0=\sqrt{n_0}\,|\,t_m\,|$
$Z_0<\alpha$ α=1.96 or 2.58
no
$\bar{t}^{\,l}=\frac{2}{n_0(n_0-1)}\sum_{i}^{n0}\sum_{j}^{n0} t_{ij}$, i, j= (C (L, k), k =1,2,3,⋯n_0)
i<j
yes
$Z_1=\frac{\sqrt{n_0(n_0-1)}\,|\,\bar{t}^{\,l}-\bar{t}_m\,|}{\sqrt{n_0+1}}$
$Z_0<\alpha$ α=1.96or 2.58
yes
no
x_m belongs to the Lth cluster
$n_0=n_0+1$
C (L, n_0)= m
D (L)=n_0
x_m does not belong to the Lth cluster
2


2
$S=\sum_{t=1}^{L} D(t)$
$n-s \geqq 2$
no
$L=L+1$
$n_0=1$
$C(L,n_0)=m$
$D(L)=1$
yes
$L=L+1$
$t_{kl}=\min\left(t_{ij};\ i,j=1,2,3,\cdots\cdots n,\ i \neq j \right.$
$\left. i,j \neq \left(\begin{matrix} C(t,k),\ k=1,2,3,\cdots D(t) \\ t=1,2,3,\cdots(L-1) \end{matrix} \right) \right)$
5
8
$D(t)$
$C(t,k),\ k=1,2,3,\cdots D(t)$
$t=1,2,3,\cdots L$
Hierar-chical clustering ?
no
stop
yes
$Q=Q+1$
$t_{pq}=\frac{1}{l_p l_q}\sum_{k=1}^{l_p}\sum_{j=1}^{l_q} x_{ij}$
$l_p=D(p),\ l_q=D(q)$
$x_i,\ i=1,2,\cdots l_p \in \{C(p,k),k=1,2,\cdots l_p\}$
$x_j,\ j=1,2,\cdots l_q \in \{C(q,k),k=1,2,\cdots l_q\}$
$T_1=\begin{pmatrix} 0 & t_{12} & t_{13} & \cdots & t_{1L} \\ t_{21} & 0 & t_{23} & & t_{2L} \\ \vdots & \vdots & & & \vdots \\ t_{L1} & t_{L2} & \cdots\cdots & & 0 \end{pmatrix}$
7

附　録(2)

軸の貢献量が等しくなるような軸の回転

回転前の任意の2軸に対する個体の座標

$$A_{ij}=\begin{pmatrix} a_{1i} & a_{1j} \\ a_{2i} & a_{2j} \\ a_{3i} & a_{3j} \\ \vdots & \vdots \\ a_{ni} & a_{nj} \end{pmatrix} \tag{1}$$

回転後の個体の座標

$$B_{ij}=\begin{pmatrix} b_{1i} & b_{1j} \\ b_{2i} & b_{2j} \\ b_{3i} & b_{3j} \\ \vdots & \vdots \\ b_{ni} & b_{nj} \end{pmatrix} \tag{2}$$

A_{ij}とB_{ij}の間には，回転角をθとすれば，

$$B_{ij}=A_{ij}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{3}$$

の関係が成立する。すなわち，

$b_{ki}=a_{ki}\cos\theta+a_{kj}\sin\theta$

$b_{kj}=-a_{ki}\sin\theta+a_{kj}\cos\theta$,

$k=1,\ 2,\ 3,\ \cdots\cdots n$

I軸の貢献量は

$$c_i=\sum_{k=1}^{n} b_{ki}^2=\sum_{k=1}^{n}(a_{ki}\cos\theta+a_{kj}\sin\theta)^2$$

$$=\cos^2\theta\sum_{k=1}^{n} a_{ki}^2+\sin^2\theta\sum_{k=1}^{n} a_{kj}^2$$

$$+2\sin\theta\cos\theta\sum_{k=1}^{n} a_{ki}a_{kj}$$

同様J軸の貢献量は

$$c_j=\sin^2\theta\sum_{k=1}^{n} a_{ki}^2+\cos^2\theta\sum_{k=1}^{n} a_{kj}^2$$

$$-2\sin\theta\cos\theta\sum_{k=1}^{n} a_{ki}a_{kj}$$

であるから，$c_i=c_j$とおくと，

$$\cos^2\theta\sum_{k=1}^{n} a_{ki}^2+\sin^2\theta\sum_{k=1}^{n} a^2{}_{kj}$$

$$+2\sin\theta\cos\theta\sum_{k=1}^{n} a_{ki}a_{kj}$$

$$=\sin^2\theta\sum_{k=1}^{n} a_{ki}^2+\cos^2\theta\sum_{k=1}^{n} a^2{}_{kj}$$

$$-2\sin\theta\cos\theta\sum_{k=1}^{n} a_{ki}a_{kj}$$

となる。整とんして，

$$(\cos^2\theta-\sin^2\theta)\sum_{k=1}^{n} a_{ki}^2-(\cos^2\theta-\sin^2\theta)\sum_{k=1}^{n} a_{kj}^2$$

$$+4\sin\theta\cos\theta\sum_{k=1}^{n} a_{ki}a_{kj}=0$$

$$\cos2\theta\sum_{k=1}^{n} a_{ki}^2-\cos2\theta\sum_{k=1}^{n} a_{kj}^2+2\sin2\theta\sum_{k=1}^{n} a_{ki}a_{kj}=0$$

$$\cos2\theta\left(\sum_{k=1}^{n} a_{ki}^2-\sum_{k=1}^{n} a_{kj}^2\right)=-2\sin2\theta\sum_{k=1}^{n} a_{ki}a_{kj}$$

$$\frac{\left(\sum_{k}^{n} a_{kj}^2-\sum_{k}^{n} a_{ki}^2\right)}{2\sum_{k=1}^{n} a_{ki}a_{kj}}=\tan2\theta$$

したがって，

$$\theta=\frac{1}{2}\tan^{-1}\left(\frac{\left(\sum_{k}^{n} a_{kj}^2-\sum_{k}^{n} a_{ki}^2\right)}{2\sum_{k}^{n} a_{ki}a_{kj}}\right)$$

である。

ここで，

$-90^\circ\leqq 2\theta\leqq 90^\circ$

とすれば，

$-45^\circ\leqq\theta\leqq 45^\circ$

であるから，

$\tan2\theta<0$ であれば $\theta<0$

$\tan2\theta>0$ であれば $\theta>0$ とする。