筑波大学体育科学系紀要　Bull. Health & Sports Sciences, Univ. of Tsukuba　12:193—213,1989.

類似度（相関係数）の統計的有意性を利用したクラスター分析手法の工夫

松　浦　義　行

Development of a procedure of cluster analysis using statistical significance test of similarity; correlation

Yoshiyuki Matsuura

This study attempted to develop a new procedure of cluster analysis with the intercorrelation matrix given. For determining whether any given individual belongs to a certain cluster, the significance test of correlation was utilized. The two individuals whose correlation was the largest and significant constituted one cluster, and other individuals which could belong to this cluster were searched with following procedures; (1), searching the individual whose mean correlation with the individuals determined already to belong to a certain cluster was the largest in the rest of individuals and significant statistically, and (2), testing non-significance of the difference between the mean intercorrelation among the individuals belonged to this cluster and the mean correlation computed at the step (1).

This clustering procedure was extended to the hierarchical clustering through defining the correlation between any two clusters with the mean correlation between the individuals belonging to one cluster and the ones to another cluster. Then, taking the order of determination for the individual to belong to a given cluster into consideration, hierarchical clustering could be accomplished. Lastly, the example was shown by the (31x31) correlation matrix computed with 31 motor ability variables.

Key words：クラスター分析，類似性，相関係数，統計的仮説検定，運動能力

Ⅰ．序　論

クラスター分析は，X_1，X_2，X_3.... X_NのN個の個体が与えられた時，このNこの個体間の類似性，または非類似性を手がかりとして，Nこの個体をmこ（m≦N）の群に分類することを目的とした統計的方法の一つである。

この分析法は個体間の類似度，または非類似度の与え方，及びアルゴリズムによって分類することが出来る。非類似度を手がかりとする場合，一般に個体間の距離が用いられるが，この距離推定のアルゴリズムには，ユークッリドの距離，重み付けユークッリド距離，標準ユークッリド距離からマハラノビスの汎距離にいたるまで種種の工夫が報告されている。また，類似度を手がかりとする場合，一般に個体間の相関係数が用いられるが，この相関係数推定のアルゴリズムには，積和，ピアソンの相関係数，順位相関係数，四分相関係数等の種種の相関係数算出の方法がある。ピアソンの相関係数で与えられる場合には，因子分析も個体分類に役立つ方法の１つである。

また，クラスター分析のアルゴリズムから分類すれば，階層的手法，非階層的手法の２つに大別できる。因子分析等は非階層的手法の１つと考えられるであろう。これに対し，階層的手法は類似度の最も大なる２個体から出発して，群と個体，群と群との類似度を定義しつつ，逐次クラスターを構成していく方法である。非階層的手法はある種の評価基準にもとづいて，個体のあるクラス

ターへの所属を判断させながら，クラスターを構成していくものである。非階層的手法ではクラスター所属の判断基準によって結果されるクラスターは異なってくる。また，階層的手法の場合は，いかなる階層で，分析を打ち切るかによって，結果されるクラスターは異なってくる。つまり，系統分類を考える場合には，階層的手法が大いに役立つが，単に分類のみを問題にする場合は非階層的手法が都合がよい。

従来，発表され，利用されて来ているクラスター分析手法の多くは，クラスター所属の判断基準に必ずしも理論的根拠が十分ではない。それは，距離（非類似度）を手がかりとする場合には，距離の確率分布が明確にされていない事によるのであろう。相関係数（類似度）を用いた場合には，その分布が判明している場合が多いのであるが，何故かその確率分布関数を利用して判断基準を設定している手法がないようである。これは，恐らく，相関係数でNこの個体間の類似度が与えられている場合は，因子分析等の手法が分類に用いられて来た事によるのではないかと思われる。しかし，単に分類のみを問題にする場合，資料として相関係数が与えられている時には，相関係数の有意性，相関係数の差異の有意性を手がかりとして考察をすすめる事のほうが統計的推測の立場からは好しい場合がある。つまり，多くの因子分析的手法は最尤度因子分析法を除いて，与えられた相関行列を母相関行列と仮定して因子の抽出が行なわれる。この仮定を必要とせず，類似性の判断基準を相関係数の差異の有意性において，個体または，変量の分類を行う方法の開発を目的として本研究は行なわれた。

II．クラスター所属の判定基準

これまで開発されて来たクラスター分析諸方法の多くは非類似性（距離）を手がかりとしている。この場合，個体相互間の距離，クラスターと個体との距離，クラスター相互間の距離を定義し，これ等の距離の大小の比較から，逐次クラスターを構成しなおしながら，最後に1つのクラスターに到達するというのが階層的クラスター分析の考え方である。非階層的クラスター分析では，距離について，ある一定の基準を設定し，この基準より小であればクラスターに所属すると判断し，大であれば所属しないと判断する。したがって，この場合，個体と個体，個体とクラスターとの2種類の距離が定義されれば十分であろう。

一方，類似性（相関係数等）を手がかりとする方法には，所属係数（coefficient of belonging-ness, B-coefficient）がよく用いられてきた。これは，あるクラスターに属する個体相互間の相関係数の平均とクラスターに属さない個体相互間の相関係数の平均との比の値とその変化量を判断の尺度としている。つまり，ある個体をあるクラスターに加えた場合のB—係数がある基準値以上であり，かつ加えない場合のB—係数との差異がある一定値以下である時，その個体は当該クラスターに所属すると判断される。

以上の諸方法の共通の弱点は，与えられた個体がいくつのクラスターに分類されるのかの判断基準が理論的根拠をもっていない点である。すなわち，どれほどの距離である時，所属すると判断するのか，どれほどのB—係数の値であって，どれほどのB—係数の変化である時，所属すると判断するのかに就いての統計的根拠がなく，これまでは経験的値が用いられているのが実情である。しかし，この方法を用いる研究の目的，問題，仮説に都合の良いように基準を設定できる柔軟性がない訳ではない。

Nこの個体を$X_1, X_2, X_3 \ldots\ldots X_N$とし，これら相互間の相関係数を$r_{ij}$；i，j＝1，2，3，…. N，$i \neq j$，とする。相互相関行列を，

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1N} \\ r_{21} & 1 & r_{23} & \cdots & r_{2N} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{N1} & r_{N2} & r_{N3} & \cdots & 1 \end{pmatrix} \qquad (1)$$

とする。Nこの個体のうち，$n_t = \sum_{i=1}^{m} n_i$

はすでにmこのクラスターに分類されているとする。すなはち，mこのクラスターの個体をそれぞれ，

第1クラスター；$x_1^1\ x_2^1\ x_3^1 \cdots x_{n_1}^1$

第2クラスター；$x_1^2\ x_2^2\ x_3^2 \cdots x_{n_2}^2$

・

・

・

第mクラスター；$x_1^m\ x_2^m\ x_3^m \cdots x_{n_m}^m$

そして，$(N-\sum_{i=1}^{m} n_i)$ この所属未決定の個体の所属を決定することを考える。すでに，第1から第（m－1）番目クラスターへの個体所属の決定は終り，第m番目のクラスターへの個体所属を検討中であるとする。

ここで，$(N-\sum_{i=1}^{m} n_i)$ この所属未決定の個体を，

$x'_1\ x'_2\ x'_3\ x'_4 \cdots x'_{n_0}$；$n_0=N-\sum_{i=1}^{m} n_i$

とする。

m番目クラスターの個体（x_i^m；i＝1，2，3..... n_m）と所属未決定個体（x'_j；j＝1，2，3..... n_0との相関係数を（$r_{j1}^m\ r_{j2}^m\ r_{j3}^m \cdots r_{jn_m}^m$）の$n_m$ことする。これらの相関係数の平均値を$\bar{r}_j^m$とすれば，$\bar{r}_j^m$は$n_0$こある。この（$\bar{r}_j^m$；j＝1，2，3..... n_0）最大値が$\bar{r}_k^m$であったとすれば，x'_kが第m番目クラスター所属の候補となる。ついで，この$\bar{r}_k^m$が次の2つの条件を満足する時，第m番目への所属が決定される。

a．$\bar{r}_k^m$が統計的に有意であること，

b．（x_i^m；i＝1，2，3,..... n_m）相互間のすべての相関係数の平均値と$\bar{r}_k^m$との間に有意な差がないこと。

aの検定は次のようにして行うことが出来る。標本数を適当に大とすれば（実際は≧40で十分），相関係数r_i^mのz-変換値をz_i^mとすれば，z_i^mは正規分布N｛$\bar{z}_\rho^m$，1/(M－3)｝；Mは標本数，に近似的に従う。したがって，（x_i^m；i＝1，2,... n_m）相互間のすべての相関係数のz－変換値の平均値は近似的に次の正規分布に従う。

$$N\left[\bar{z}_\rho^m,\ 2/(n_m(n_m-1)(M-3))\right] \qquad (2)$$

これは次のようにして，導かれる。

$\bar{z}_i^m$はMが十分大なる時には，N｛$\bar{z}_\rho^m$，1/(M－3)｝に従うと仮定出来る。iは（$x^m{}_1\ x^m{}_2\ x^m{}_3 \cdots x^m_{n_m}$）の相互間の対の数だけ変化するから，i＝1，2，3,..... $n_m(n_m-1)/2$，である。したがって，

$n_m(n_m-1)/2=l$とすれば，$\sum_{i=1}^{l} z_i^m$は，

$$N\left[l\bar{z}_\rho^m,\ l/(M-3)\right]$$

に従う。ついで，$\sum_{i=1}^{l} z_i^m/l$は，

$$N\left[\bar{z}_\rho^m,\ l/\{l^2(M-3)\}\right]$$
$$=N\left[\bar{z}_\rho^m,\ 1/\{l(M-3)\}\right] \quad *1,2$$

に従う。したがって，$l=n_m(n_m-1)/2$を代入すれば，$\bar{z}^m$（$=\sum_{i=1}^{l} z_i/l$）は

$$N\left[\bar{z}_\rho^m,\ 2/\{n_m(n_m-1)(M-3)\}\right]$$

に従う。

x_i^m（i＝1，2，3,..... n_m）とx'_j（j＝1，2，3,..... n_0）の任意の1つx'_kとのn_mこの相関係数（r_{jk}^m；j＝1，2，3,..... n_m）をz－変換した値（z_{jk}^m；j＝1，2，3,..... n_m）の平均値$\bar{z}_k^m$の平均値は，(2)と同様に，

$$N\left[\bar{z}_{\rho k}^m,\ 1/\{n_m(M-3)\}\right] \qquad (3)$$

に従う。

それ故，$(\bar{Z}^m-\bar{Z}_k^m)$は，

$$N\left[\bar{z}_\rho^m-\bar{z}_{\rho k}^m,\ \frac{2}{n_m(n_m-1)(M-3)}+\frac{1}{n_m(M-3)}\right] \qquad (4)$$

に従う。

したがって，

$$z_0=\frac{[(\bar{z}^m-\bar{z}_k^m)-(\bar{z}_\rho^m-\bar{z}_{\rho k}^m)]}{\sqrt{\frac{2}{n_m(n_m-1)(M-3)}+\frac{1}{n_m(M-3)}}} \qquad (5)$$

はN（0，1）に従うことになる。

ここで，帰無仮説を

$\bar{z}_\rho^m-\bar{z}_{\rho k}^m=0$

とすれば，(5)は

$$z_0=\frac{(\bar{z}^m-\bar{z}_k^m)}{\sqrt{\frac{2}{n_m(n_m-1)(M-3)}+\frac{1}{n_m(M-3)}}} \qquad (6)$$

となり，(6)式が第mクラスター内個体相互間相関係数の平均値と，個体xと第mクラスラー内個体とのn_mこの相関係数の平均値との差異の有意性の検定に役立つ式である。(6)式を両側検定に用いる場合には，

$$z_0=\frac{|\bar{z}^m-\bar{z}_k^m|}{\sqrt{\frac{2}{n_m(n_m-1)(M-3)}+\frac{1}{n_m(M-3)}}} \qquad (7)$$

となる。

さて，所属未決定の個体は$x'_1\ x'_2\ x'_3 \cdots x'_{n_0}$と$n_0$こあるから，これら$n_0$この個体のすべてに就いて，第mクラスター所属個体との相関係数の平均値を，

$(\bar{r}'_1\ \bar{r}'_2\ \bar{r}'_3 \cdots \bar{r}'_{n_0})$

とする時，このn_0この相関係数の最大値を，

$\bar{r}'_k=\max(\bar{r}'_1\ \bar{r}'_2\ \bar{r}'_3 \cdots \bar{r}'_{n_0})$

とし，この$\bar{r}'_k$が有意であり，かつ(7)式での平均相関係数との差が有意でない場合に，x'_kは第mクラスターに所属すると判断する。

$\bar{r}'_k$の有意性の検定には，

$$t=\frac{\bar{r}'_k\sqrt{M-2}}{\sqrt{1-\bar{r}'^2_k}} \tag{8}$$

が，自由度（M－2）のt－分布に従うことを利用する事が出来る。しかし，(8)式は，Mが適当に大なる時には，N（0，1）に近似的に従うことがわかっている。したがって，Mが適当に大なる時には，$t\geqq1.96$；$\alpha=0.05$，$t\geqq2.58$；$\alpha=0.01$を基準として用いることが出来るので，電子計算機のプログラムには便利である。この判定をさらに厳格に次のように行うことも出来る。

任意の相関係数r_lに対応するtの値をt_lとする。

$$t_l=\frac{r_l\sqrt{M-2}}{\sqrt{1-r_l^2}} \tag{9}$$

Mが適当に大なる時は，tはN（0，1）に従うから，mこのtの値の平均値の分布は，N（0，1/m)に従うことは容易に導かれる。*2) したがって，

$$\bar{t}=\frac{\sqrt{M-2}}{m}\sum_{i=1}^{m}\frac{r_i}{\sqrt{1-r_i^2}} \tag{10}$$

とし，

$$t_0=\frac{\bar{t}}{\frac{1}{m}}=m\bar{t} \tag{11}$$

を，1.96または2.58と比較することによって，平均相関係数の有意性を検討することが出来る。ここで，m＝1の場合（第1クラスター要素の探索の場合，m＝1とすれば，明らかなように）(9)式が検定のための統計値となる。

III．アルゴリズム

次のアルゴリズムに従って計算を進めることによって，非階層的クラスター分析が達成される。

(1) 相関行列の要素の中から絶対値が最大である相関係数を見付ける。それをr_{kl}とする。

(2) r_{kl}が有意であることを検討する。それには，

$$t_0=\frac{r_{kl}\sqrt{M-2}}{\sqrt{1-r_{kl}^2}}\qquad M；標本数$$

を求め，Mが十分大である時には，

$t_0\geqq1.96$または，$t_0\geqq2.58$ (12)

によって，そうでない場合には，

$t_0\geqq t(df=M-2,\alpha=0.05\text{ or }\alpha=0.01)$ (13)

によって，r_{kl}の有意性を確かめる。

(3) (12)または(13)を満足すればr_{kl}は有意であるから，x_kとx_lがクラスターを構成すると判断する。

(4) x_kとx_l以外の個体のx_kとx_lとの相関係数を相関行列から見出し，それらをz－変換する。このz－変換値の平均値をx_k，x_l以外のすべての（N－2）この個体について求め，それらを今$(\bar{z}_i；i=1，2，3,\ldots..\ n，i\neq k，l)$とする時，$\bar{z}_i$の最大値を見付け，

$\bar{z}_m=\max(\bar{z}_i；i=1,2,3\ldots..\ n, i\neq k,l)$

とする。かつ，これに対する相関係数を$\bar{r}_m$とする。

(5) $\bar{r}_m$の有意性を検討する。有意であれば，r_{kl}と$\bar{r}_m$（$\bar{z}_m$に対応する相関係数）の差異の有意性を検定する。これには，$n_m=2$として(7)式を用いる。すなはち，

$$z_0=\frac{|z_{kl}-\bar{z}_m|}{\sqrt{\frac{1}{(M-3)}+\frac{1}{2(M-3)}}} \tag{14}$$

を用いればよい。

$z_0<1.96$または$z_0<2.58$

であれば，x_mはx_k，x_lと共に同一クラスターを構成すると判断する。

もし，$\bar{r}_m$が有意でない場合には，x_k，x_lが構成するクラスターにはx_mは属さないと判断し，このクラスターに属する個体の探索を打ち切る。

(6) x_mが（x_k，x_l）のクラスターに属すると判断された場合には，（x_k，x_l，x_m)以外の（N－3）この個体の中から，このクラスターに所属する個体を探索する。これには，(4)，(5)の計算を$n_m=3$の場合について行えばよい。

(7) したがって，(4)，(5)の計算手続を，前節で述べた，a，bの条件のいずれかが否定される迄繰り返す。したがって，a，bのいずれかが否定された時，検討中のクラスターの所属個体の探索を打ち切り，ここまでに所属すると判断された個体，

（x_k x_l x_m・・・・・）

が1つのクラスターを構成すると判断する。

(8) 1つのクラスターは見出されたから，次に別のクラスターの構成に進む。まず，すでにクラスターの所属が判明した個体を除いた，他の個体

相互の相関係数のうち最大の値を示す2個体を見付ける。今これらを x_h^1, x_l^1 とし，相関係数を r とする。以下，(2)から(7)までの手続を繰り返せばよい。

(9) 以上の手続きを繰り返し，すべての個体がいずれかのクラスターに所属することの判断がなされるまで行えばよい。

以上のアルゴリズムをフローチャートで示せば付録1の通りである。

IV. 階層的クラスター分析手法への拡大

前節で，与えられたNこの個体は，mこのクラスターに分類された。今これらのクラスターの要素である個体をそれぞれ次の通りとする。

第1クラスター；$(x_1^1\ x_2^1\ x_3^1 \cdots\cdots x_{n_1}^1)$

第2クラスター；$(x_1^2\ x_2^2\ x_3^2 \cdots\cdots x_{n_2}^2)$

第3クラスター；$(x_1^3\ x_2^3\ x_3^3 \cdots\cdots x_{n_3}^3)$

・

・

・

第mクラスター；$(x_1^m\ x_2^m\ x_3^m \cdots\cdots x_{n_m}^m)$

ここで，

$$N=\sum_{i=1}^{m} n_i \qquad (15)$$

である。

クラスター間の相関係数を次のように定義する。

任意の2つのクラスターとその要素を，C_i $(x_1^i\ x_2^i\ x_3^i \cdots\cdots x_{ni}^i)$，$C_j$ $(x_1^j\ x_2^j\ x_3^j \cdots\cdots x_{nj}^j)$ とする。ついで，x_k^i と x_l^j との相関係数を r_{kl}^{ij} とすれば，r_{kl}^{ij} は C_i, C_j のすべての要素間で考えれば，$n_i \times n_j$ こある。この $n_i \times n_j$ この相関係数の平均値をもって，クラスター C_i と C_j との相関係数とする。つまり，z_{kl}^{ij} を r_{kl}^{ij} の z 一変換値とすれば，

$$\bar{z}^{ij}=\sum_{k=1}^{n_i}\sum_{l=1}^{n_j} z_{kl}^{ij} / (n_i\ n_j) \qquad (16)$$

$\bar{z}^{ij}$ を相関係数に逆変換したものを $\bar{r}^{ij}$ とすれば，これが求める平均相関係数である。

このアルゴリズムに従って，クラスター相互間の相関係数をすべて求め，行列に整理したものを R とする。すなわち，

$$R=\begin{pmatrix} 1 & \bar{r}^{12} & \bar{r}^{13} & \cdots & \bar{r}^{1m} \\ \bar{r}^{21} & 1 & \bar{r}^{23} & \cdots & \bar{r}^{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ \bar{r}^{m1} & \bar{r}^{m2} & \bar{r}^{m3} & \cdots & 1 \end{pmatrix} \qquad (17)$$

をクラスター相互間の相関行列と定義する。

この(17)の相関行列を資料として，前節のアルゴリズムを実施すれば，第2段階のクラスター分析が完了する。すなわち，クラスターのクラスターが構成されることになる。

この手続をクラスターが1つになるまで，または，残ったクラスター間の相関係数が，II節のaあるいはbのいずれかの条件を満足しない情況となるまで繰り返せばよい。

このアルゴリズムに従った階層的クラスター分析では，従来の距離を資料とした階層的クラスター分析手法によって得られる階層の数より少ないのが一般である。また，最終的に1つのクラスターにまとまるとは限らない。それは，クラスター所属の決定に確率的判定基準を用いているからである。

さて，非階層的クラスター分析に於て，前節までに述べたように，任意のクラスターに属する要素と判断される順位を考慮すると，すでに階層をなしている。すなはち，i番目クラスターまでの要素は，まず(i－1)番目クラスターまでの(i－1) このクラスターの要素以外の個体の中で，絶対値が最大で，かつ有意な相関係数を示す2個体を x_1^i, x_2^i として，i番目のクラスターの核要素 (core element) とする。ついで，残りの個体の中で，この x_1^i, x_2^i との相関係数の平均値が最大で，かつ有意である個体を x_3^i とし x_1^i と，x_2^i との相関係数とこの平均相係数の差が有意でない場合，x_3^i を $(x_1^i,\ x_2^i)$ のクラスターの要素であると判断し，$(x_1^i,\ x_2^i,\ x_3^i)$ とクラスターを再構成する。以下，この手続を上述の平均相関係数が有意でないか，クラスター要素間の相関係数の平均値と有意差を示すまでクラスター要素の探索を続ける。この過程が非階層的クラスター要素決定の過程である。したがって，クラスター要素として判断される順位を考慮すれば，階層的になっている訳である。

ここで，工夫された階層的クラスター分析手法

Table 1. Correlation matrix; 31 motor ability test items

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1. standing height																															
2. weight	66																														
3. chest girth	59	70																													
4. sitting height	53	60	60																												
5. vertieal jump	42	20	43	34																											
6. standing broad jump	29	13	30	06	52																										
7. direction change run	23	09	22	20	43	53																									
8. 100m dash	30	22	36	23	43	55	50																								
9. 50m shuttle run	27	10	18	15	36	52	37	52																							
10. baseball throw	49	34	46	39	35	37	37	43	31																						
11. 8Lbs. shot put	40	24	36	29	30	35	32	36	32	43																					
12. running broad jump	56	25	56	40	45	46	50	61	43	50	43																				
13. standing high jump	45	20	44	42	48	55	43	14	46	9	41	70																			
14. running high jump	42	28	57	45	58	66	47	65	54	50	51	89	78																		
15. push-ups	12	-08	30	13	20	24	32	31	36	30	32	27	14	29																	
16. sit-ups	22	15	36	19	32	38	37	37	43	39	32	39	30	41	53																
17. chins	12	08	30	04	35	38	35	34	39	33	29	27	20	37	40	43															
18. squat jump	21	14	31	10	32	35	20	41	41	29	23	44	33	55	52	45	42														
19. back strength	25	20	32	24	11	29	23	32	24	32	34	24	24	32	11	15	19	14													
20. leg strengh	17	16	37	15	15	37	26	29	21	34	41	39	40	47	11	12	22	25	55												
21. abdominal strength	30	24	42	19	17	33	26	36	26	35	39	35	33	48	24	24	20	30	47	49											
22. grip (R) strength	20	13	43	21	20	39	28	43	25	32	55	42	40	56	27	34	27	35	48	44	59										
23. grip (L) strength	17	12	39	22	08	20	19	30	20	29	42	29	24	37	14	22	21	29	43	32	34	81									
24. squat thrust	23	17	-07	25	26	29	25	25	22	27	30	30	19	22	31	38	26	32	15	24	15	25	16								
25. side step	15	-03	-09	14	30	25	31	24	21	22	31	33	20	19	32	37	22	30	17	30	14	33	30	44							
26. foot & toe balance	10	03	19	10	22	37	29	40	36	27	29	39	36	41	13	17	25	29	15	25	24	25	22	08	15						
27. frog balance	09	06	24	09	13	26	34	37	35	32	23	38	40	46	15	14	23	21	17	15	26	24	26	07	10	54					
28. breath holding	19	20	26	20	04	01	04	10	01	09	10	16	09	10	01	09	02	14	04	04	04	10	03	13	12	20	-04				
29. drop-of	04	15	13	43	07	13	04	09	05	10	11	12	13	24	03	12	01	09	05	05	02	15	02	10	07	12	13	52			
30. trunk flexion	23	04	14	21	13	24	25	20	13	06	05	23	13	32	04	04	05	16	01	12	15	14	13	11	15	03	02	15	10		
31. trunk extension	19	05	12	16	10	13	24	22	10	10	09	31	16	31	09	07	03	18	03	04	14	12	15	12	20	06	08	03	09	53	

decimal point omitted.

は，前述のように個体をクラスターに分類し，分類されたクラスターを個体として更にクラスターに分類し，再び前段階で得られたクラスター（最初の段階から見ればクラスターのクラスター）を個体として分類するという様に，非階層的クラスター分析を，次第に次数の減少していく相関行列を資料として繰り返していくのである。すでに，述べたように，各段階において，クラスターの要素の決定には順位がある。この順位は相関係数の大きさに依存している。このようにして，要素決定の順位を考慮すれば，容易に階層的クラスター構成を推測することが出来る。したがって，従来のクラスター分析手法に於けるように主観的な基準によってクラスターを決定する事も出来る便利さも有している。

Ⅴ．運動能力31変量の分類；非階層的分類

Table 1は運動能力31変量についての相関行列である。この相関行列に，ここで工夫された方法を適用した結果がTable 2に示されている。すなはち，10このクラスターに分類され，各クラスターの要素は次のとおりである。

C_1；走幅跳び，走高飛び，立高飛び，
C_2；立幅跳び，100m走，50m走，往復走，方向変換走，
C_3；握力（右），握力（左）
C_4；体重，胸囲，身長，座高
C_5；背筋力，脚力，腹筋力，8ポンド砲丸投，野球ボール投
C_6；片足つま先立ち，蛙立ち
C_7；腕立伏臥腕屈伸，上体起し，スカット・ジャンプ，懸垂
C_8；立位体前屈，立位体後屈
C_9；運動後の息こらえ，ドロップ・オフ
C_{10}；スカット・スラスト，サイド・ステップ

Table 2. Clusters and their elements produced by new procedure; 31 motor ability items

Cluster	No. of items	item
Cluster 1	3	12 14 13
Cluster 2	5	6 8 9 7 5
Cluster 3	2	22 23
Cluster 4	4	2 3 1 4
Cluster 5	5	19 20 21 11 10
Cluster 6	2	26 27
Cluster 7	4	15 16 18 17
Cluster 8	2	30 31
Cluster 9	2	28 29
Cluster 10	2	24 25

Note; Item corresponds to the item shown in the left column of table 1.

同じ相関行列に主因子解とノーマル・ベリマックス基準による直交回転をほどこして得られた多因子解を用いて，31変量を分類するとTable 3のF-procedureの欄の通りであり，17のクラスターに分類された。また，B－係数を用いた結果はTable 3のB-procedureの欄の通りで，11クラスターに分類された。

3解法の間で完全に一致するクラスターは，C_1 (12, 14, 13)，C_3 (22, 23)，C_6 (26, 27)，C_8 (30, 31)，C_9 (28, 29)，C_{10} (24, 25) の6クラスターであった。また，3解法の間で一部の一致を示すものは，

C_2；C（6，8，9，7，5），F（8，9），B（5，6，7，8，9），

C_4；C（2，3，1，4），F（1，2，3，4，10，11），B（1，2，3，4）

の2クラスターであった。

B一係数を用いた結果と比較すると，要素の完全な一致を示したクラスターは前述の6クラスターに加えて，C_2（6，8，9，7，5），C_4（2，3，1，4）とC_7 (15, 16, 18, 17) の3クラスターであった。さらに，一部の一致を示したものは，次の1クラスターであり，C_5 (19, 20, 21, 11, 10) がB一係数による結果のB (19, 20, 21)，B (10, 11) クラスターに分解されて対応していた。すなはち，

C_5(19,20,21,11,10) ——┬— B (19,20,21)
　　　　　　　　　　　└— B (10,11)

であった。したがって，B一係数を用いた結果では完全な一致が9クラスター，部分的一致が1クラスターであり，すべてのクラスターがB一係数を用いた場合の結果と対応していた。これに対し，因子分析の結果から得られたクラスターは17であり，そのうち8クラスターは1個体を要素とするものであった。分類の立場からは1要素でクラスターを構成するとするのは適当とはいえない。また，1クラスターを構成する要素の数は他の2方法の場合より概して少なく，しかも，2要素から

Table 3. Comparison in the clusters and their elements between three different procedures

cluster	C-procedure	F-procedure	B-procedure
C1*	12, 14, 13	12, 13, 14	12, 13, 14
C2+	6, 8, 9, 7, 5	8, 9	5, 6, 7, 8, 9
C3*	22, 23	22, 23	22, 23
C4+	2, 3, 1, 4	1, 2, 3, 4, 10, 11	1, 2, 3, 4
C5	19, 20, 21, 11, 10		19, 20, 21
C6*	26, 27	26, 27	26, 27
C7+	15, 16, 18, 17		15, 16, 17, 18
C8*	30, 31	30, 31	30, 31
C9*	28, 29	28, 29	28, 29
C10*	24, 25	24, 25	24, 25
C11			10, 11
C12		18	
C13		5	
C14		21	
C15		16	
C16		19, 7	
C17		6	
C18		15	
C19		17	
C20		20	
total	10 clusters	17 clusters	11 Clusters

Note; These numbers correspond to the item shown in the the left column of table 1.
C-procedure stands for Cluster analysis procedure developed in this paper, F-prodecure for Factor analytic procedure; principal factor solution and Normal Varimax rotation, and B-procedure for classification procedure with B-coefficient.
* stands for the clusters commonly produced in three procedures.
+ stands for the clusters commonly produced in C-procedure and B-procedure.

成るクラスターは他の2方法の結果と完全に一致していた。したがって，因子分析を分類に応用する場合，より細く分類される傾向があると考えられる。しかし，B一係数による結果とでは，完全な一致が9クラスター，部分的一致が1クラスターとすべてのクラスター間に対応が見られた。B一係数を用いる手法はクラスター所属の判断基準の客観性は低いが，統計的推測を導入した本法と比較して大差はなく従来からクラスター分析の一方法として，また，因子分析のための変量分類の方法として用いられて来たことの有効性の一面が確かめられたと同時に，ここで工夫されたクラスター分析手法の有効性が示されたとも考えられる。

VI. 運動能力31変量の分類；階層的分類

V. で得られた10クラスター相互の相関係数はTable 4の通りである。この相関行列に再びV. におけると同様なクラスター分析を施した結果が，Table 5である。つまり，5で得られた10このクラスターは，C_1, C_4, C_6, C_5, C_2, C_7, C_3の順位でクラスター1を構成し，C_8, C_9, C_{10}はそれぞれ単独に別のクラスターを構成すると考えられた。したがって，第2段階のクラスター分析で4クラスターが得られた。この4クラスター相互間の相関係数はTable 6の通りである。この相関行列が示すように，クラスター相互の相関係数は極めて低い値となっている。この相関行列に再びクラスター分析をほどこすと，Table 7の通り3クラス

Table 4. Inter-cluster correlation matrix for clustering of stage 2.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	1									
C2	.385	1								
C3	.423	.237	1							
C4	.502	.254	.238	1						
C5	.400	.423	.302	.294	1					
C6	.400	.242	.113	.310	.233	1				
C7	.334	.262	.170	.341	.238	.196	1			
C8	.244	.135	.143	.174	.079	.047	.082	1		
C9	.140	.075	.202	.058	.064	.100	.063	.092	1	
C10	.239	.261	.095	.258	.226	.100	.310	.145	.105	1

Table 5. Clusters and their elements produced by new procedure at the 2nd stage; 10 clusters

Cluster	No. of items	Cluster #
Cluster 1	7	C1 C4 C6 C5 C2 C7 C3
Cluster 2	1	C8
Cluster 3	1	C9
Cluster 4	1	C10

Note ; Cluster# ; Ci corresponds to the cluster# shown in the left column of table 3.

Table 6. Inter-clustercorrelation matrix for clustering of stage 3.

	C′1	C′2	C′3	C′4
C′1	1			
C′2	.130	1		
C′3	.101	.092	1	
C′4	.214	.145	.105	1

Table 7. Cluster and its elements produced by new procedure at the 3rd stage; 4 clusters

Cluster	No. of items	Cluster #
Cluster 1	2	C′1 C′4
Cluster 2	1	C′2
Cluster 3	1	C′3

Note ; Cluster# ; C′i corresponds to the cluster# shown in the left column of table 5.

Table 8. Inter-cluster correlation matrix for clustering of state 4.

	C″1	C″2	C″3
C″1	1		
C″2	.137	1	
C″3	.103	.092	1

Table 9. Cluster and its elements produced by new procedure at the 4th stage; 3 clusters

Cluster	No. of items	Cluster #
Cluster 1	1	C″1
Cluster 2	1	C″2
Cluster 3	1	C″3

Note; Cluster # ; C″i, corresponds to the cluster# shown in the left column of table 7.

ターが構成された。しかし，クラスター2，3はそれぞれ1つのクラスターを要素とするものであり，第1段階の分析で構成されたクラスター8，9は第2段階でも他のクラスターと融合してクラスターを構成することはなく，第3段階でも他のクラスターと融合することはなかった。ついで，第3段階において得られた3クラスター相互間の相関行列はTable 8の通りであり，これを再び同じ手続でクラスター分析をほどこした結果がTable 9の通り，いずれの要素も他と融合してクラスターを構成しないことがわかる。したがって，第4段階の分析で，この手続を停止して，以上の過程をデンドログラムに描くと，Fig. 1の通りである。さらに，相関係数の大きさを手がかりにクラスター所属をその大きさの順に加えていくという手続の性質を考慮して，再び以上の過程をデンドログラムにあらわしたものが，Fig. 2である。Fig. 2において，C_8 (30, 31)，C_9 (28, 29) の2クラスターが他のクラスターと融合しないのは，Table 8が示すように，この段階での3つのクラスター相互間の相関係数が有意でないことによるものである。

そこで，従来の階層的クラスター分析のように，最後に1つのクラスターに融合するようにするため，第3段階での分析で，相関係数の有意性の検定の手続を除くとFig. 3のデンドログラムが得られる。Fig. 2及びFig. 3においては，従来のクラスター分析に於けるように主観的にクラスターを同定することが出来る。

Ⅶ. 討論まとめ

個体のクラスターへの所属決定の判断に統計的仮説検定の手続を用いることによって，従来距離を用いたクラスター分析において広く用いられて来た主観的判断を避ける工夫をした。しかし，本方法を応用するには，個体間の類似性をあらわす統計値の分布がわかっている必要がある。この点は，判断を客観的なものにするが，それだけ適用の範囲を狭めることになる。B一係数を用いた結果とは非常に良い一致を示しており，客観的基準を利用している点かから本法の分類結果の一義性が保証され得るであろう。しかし，この一義性は標本の大きさによって，相関係数の有意水準が異なることから再びくずれてしまうことになる。しかし，因子分析におけるような因子モデルの差異

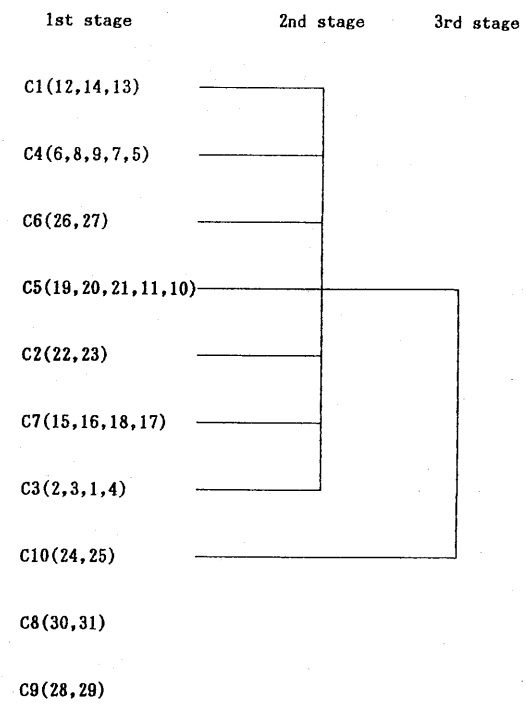


Fig. 1 Hierarchical constructs of 31 motor ability test items--I

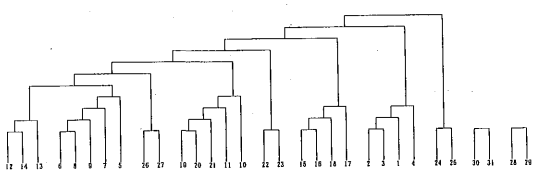


Fig. 2 Hierarchical construct of 31 motor ability test items with testing procedure of significance of mean correlation.

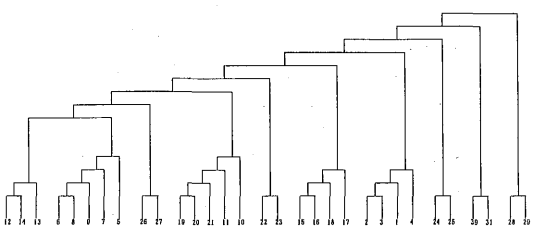


Fig. 3 Hierarchical construct of 31 motor ability test items without testing procedure of significance of mean correlation at the last stage of clustering.

によって因子解が異なるというような不確定性ではなく，前にも述べたように標本の大きさによるものであり，統計的推測の基本的条件にかかわるものである。これは統計的推測の立場をとる限り避け得ないものといえよう。

Gower (1967) は，クラスター分析の諸方法について，比較検討しているが，すべて非類似性測度から出発する方法を取り上げており，ここで工夫された方法と比較する事は出来ないが，統計的推測の論理を用いない事は，distribution-free の特質を方法論に保持出来る事から，その応用が広いと述べている。この点に就いては本方法は，すでに述べたように，適用に制限がある事は短所といわねばならない。しかし，方法の客観性を高めようとすれば，高める為に或る条件が付加される事は避け得ないであろう。さらに，Anderberg (1973)，Everitt (1974) 及び Hartingen (1975) は非類似性（距離），類似性（相関係数等）を資料としたクラスター分析の方法について総括的にまとめて論じているが，統計的仮説検定の手法を用いる方法につては論じていない。しかし，Cormack (1971) はクラスター分析の諸方法が必しも適切な分類を最終的結論として与えるとは言えないと否定的論述をしており，この方法によって得られる分類は以後の研究のための手がかりとしての意味をもつがそれ以上の有効性を持たせるのは危険であると述べ，この方法の弱点を指摘している。しかし，個体の分類を，以後の研究を進める上で探索的手法としてクラスター分析の手法は有用なものいえるであろう。そうであるが故に，Cormack (1971) の批判にかかわらず，いまでも探索的意味で多用されていると考えられる。

相関係数をデータとするのであれば，因子分析が十分変量分類に役立つとも言えるが，因子分析の結果と比較すれば，分類という問題に限れば本方法のほうがより簡潔な結果を与え得ると考えられる。更に，階層的クラスター解を得ておけば，従来の方法と同様に，利用者の主観的判断も可能となり，クラスター分析の一つの長所を活かす事が出来る。

引用，参考文献

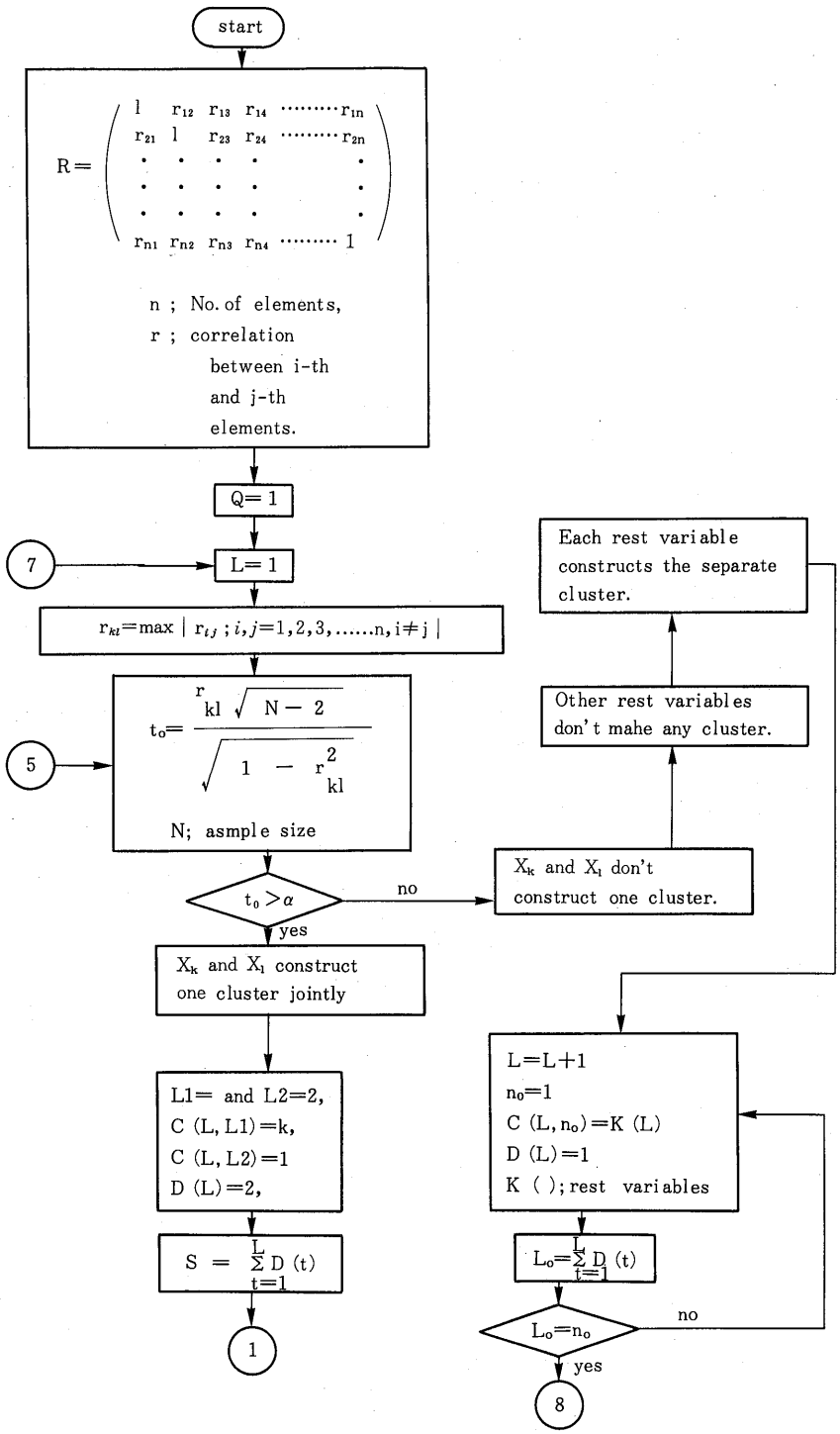
1) Anderberg, M.R., Cluster analysis for application, Academic Press, 1973
2) Chatfield, C. and Collins, A.J., Introduction to multivariate analysis, Chapman & Hall Ltd, 1984.
3) Everitt, B.S., Cluster analysis, London, Heineman, 1974
4) Everitt, B.S., Unresolved problems in cluster analysis, Biometics, 35, pp.169-182, 1979
5) Fiedman, H.P. and Rubin, J., On some invariant criteria for grouping data, J. of American Statistical Association, 62, pp. 1159-1178, 1967.
6) Fisz, M., Mathematical statistics, 3rd ed. John Wiley & Sons, 1963
7) Gower, J.C., A comparison of some methods of cluster analysis, Biometrics, 23, pp. 623-637, 1967.
8) Harman, H.H., Modern factor analysis, PP. 28-131, The Univ. of Chicago press, 1962.
9) 松浦義行，運動能力の因子構造，PP. 170-171，不昧堂出版，1969.
10) 松浦義行，行動科学における因子分析法，PP.1 124-128，不昧堂出版，1972
11) 奥野忠一，久米　均，芳賀敏郎，吉沢　正，(1972)，多変量解析法，PP. 396-398，日科技連，1972.
12) 前掲書，PP. 400-410.
13) Wilks, S.S., Matematical statistics, P. 189, John Wiley & Sons, 1962.

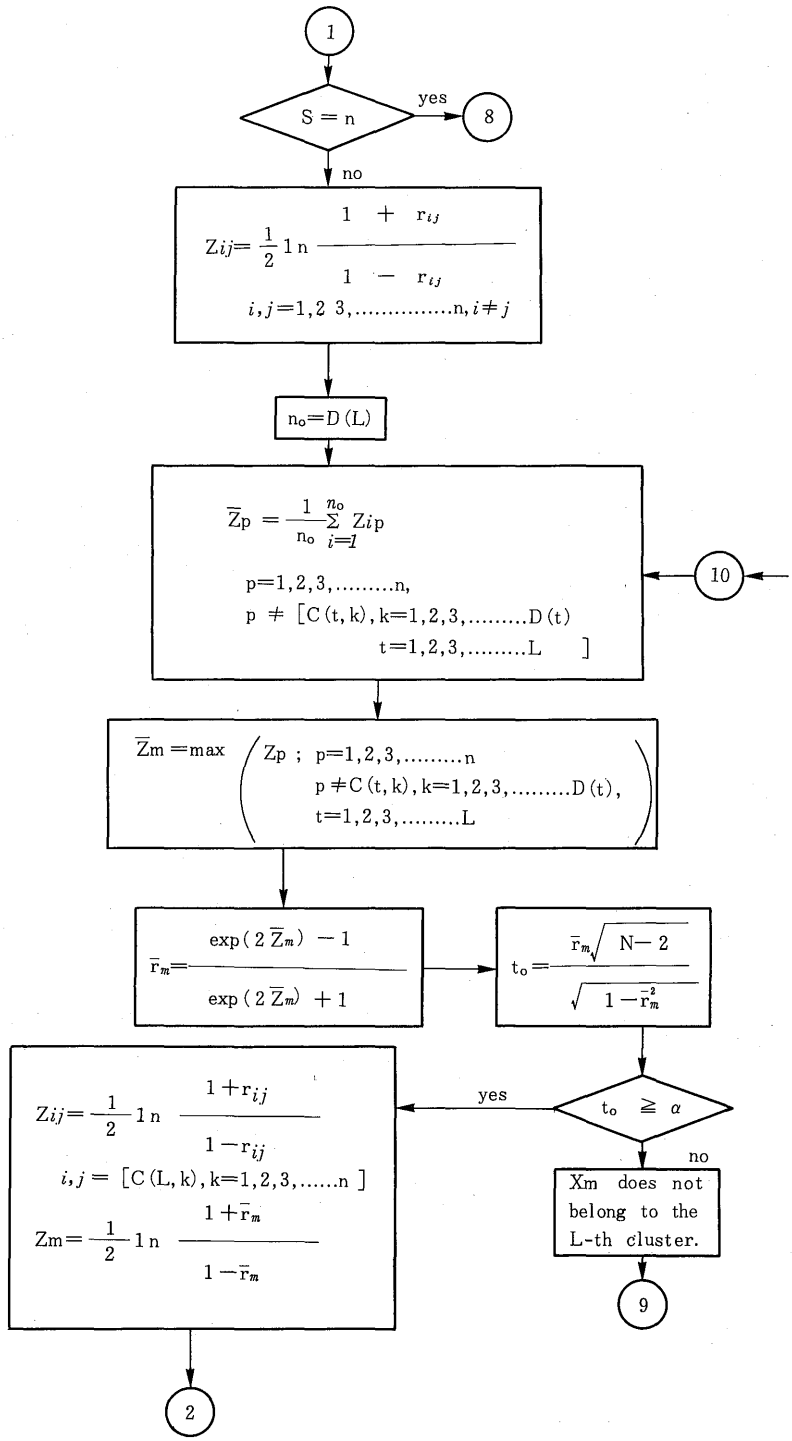
注

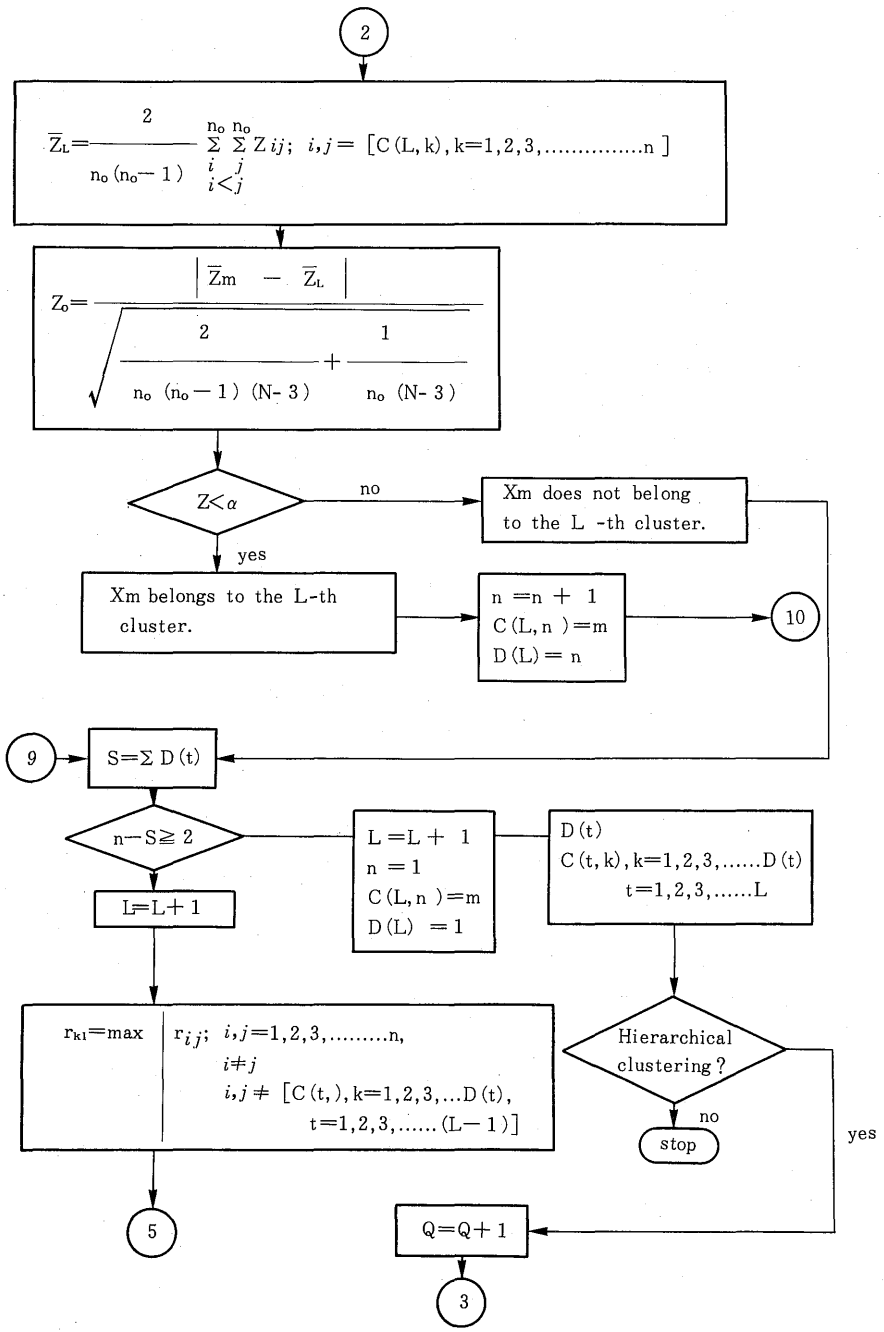
＊1) $Y=\sum_{i=1}^{n} a_i X_i$ において，X_i；i＝1，2，3，... nが独立とすれば，$\sigma_y^2=\sum_{i=1}^{n} a_i^2 \sigma_i^2$である。ただし，$\sigma_i^2$は$X_i$の分散である。

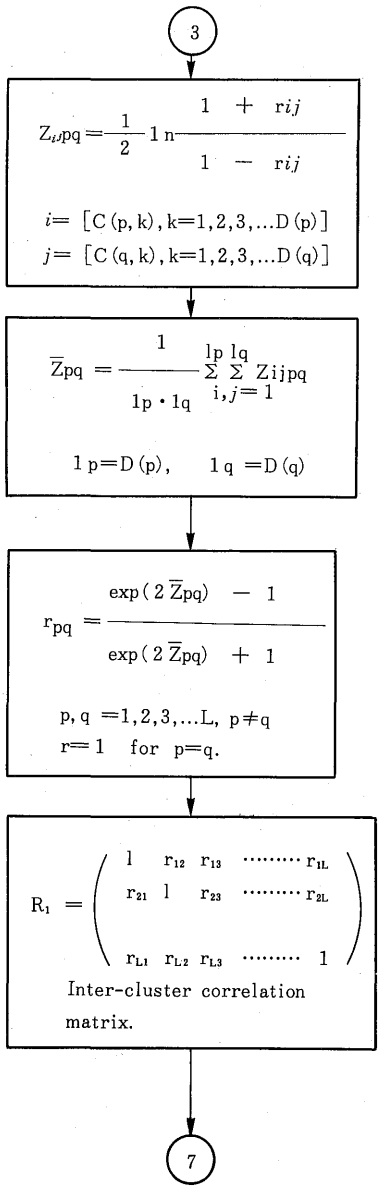
＊2) $Y=\sum_{i=1}^{m} t_i$はN（0，m）に従う。したがって，$Y=\sum_{i=1}^{m} t_i/m$ は$N(0, m/m^2)=N(0, 1/m)$に従う。

Appendix; Flow chart for computation of non-hierarchical and hierarchical clustering procedures developed.




1
S = n
yes
8
no
$Z_{ij} = \frac{1}{2} \ln \frac{1 + r_{ij}}{1 - r_{ij}}$
$i, j = 1, 2, 3, \ldots\ldots\ldots\ldots n, i \neq j$
$n_o = D(L)$
$\bar{Z}_p = \frac{1}{n_o} \sum_{i=1}^{n_o} Z_{ip}$
$p = 1, 2, 3, \ldots\ldots\ldots n,$
$p \neq [C(t,k), k = 1, 2, 3, \ldots\ldots\ldots D(t)$
$t = 1, 2, 3, \ldots\ldots\ldots L$]
10
$\bar{Z}_m = \max \left(Z_p ; p = 1, 2, 3, \ldots\ldots\ldots n, p \neq C(t,k), k = 1, 2, 3, \ldots\ldots\ldots D(t), t = 1, 2, 3, \ldots\ldots\ldots L \right)$
$\bar{r}_m = \frac{\exp(2\bar{Z}_m) - 1}{\exp(2\bar{Z}_m) + 1}$
$t_o = \frac{\bar{r}_m \sqrt{N - 2}}{\sqrt{1 - \bar{r}_m^2}}$
$t_o \geqq \alpha$
yes
no
Xm does not belong to the L-th cluster.
9
$Z_{ij} = \frac{1}{2} \ln \frac{1 + r_{ij}}{1 - r_{ij}}$
$i, j = [C(L,k), k = 1, 2, 3, \ldots\ldots n]$
$Z_m = \frac{1}{2} \ln \frac{1 + \bar{r}_m}{1 - \bar{r}_m}$
2


2
$\bar{Z}_L = \frac{2}{n_o(n_o-1)} \sum_{i<j}^{n_o} \sum_{j}^{n_o} Z_{ij}$; $i,j = [C(L,k), k=1,2,3,\ldots\ldots\ldots\ldots n]$
$Z_o = \frac{|\bar{Z}m - \bar{Z}_L|}{\sqrt{\frac{2}{n_o(n_o-1)(N-3)} + \frac{1}{n_o(N-3)}}}$
$Z<\alpha$
no
Xm does not belong to the L -th cluster.
yes
Xm belongs to the L-th cluster.
n =n + 1
C(L,n)=m
D(L)= n
10
9
S=Σ D(t)
n−S≧ 2
L =L + 1
n = 1
C(L,n)=m
D(L) = 1
D(t)
C(t,k), k=1,2,3,.....D(t)
t=1,2,3,.....L
L=L+1
r_{kl}=max r_{ij}; i,j=1,2,3,.....n, $i \neq j$
$i,j \neq$ [C(t,), k=1,2,3,...D(t), t=1,2,3,.....(L−1)]
Hierarchical clustering ?
no
stop
yes
5
Q=Q+1
3


3
$Z_{ij}pq = \frac{1}{2} \ln \frac{1 + rij}{1 - rij}$
$i = [C(p,k), k=1,2,3,\ldots D(p)]$
$j = [C(q,k), k=1,2,3,\ldots D(q)]$
$\bar{Z}pq = \frac{1}{lp \cdot lq} \sum_{i,j=1}^{lp\ lq} Zijpq$
$lp = D(p), \quad lq = D(q)$
$r_{pq} = \frac{\exp(2\bar{Z}pq) - 1}{\exp(2\bar{Z}pq) + 1}$
$p, q = 1,2,3,\ldots L, \ p \neq q$
$r = 1$ for $p = q$.
$R_1 = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots\cdots & r_{1L} \\ r_{21} & 1 & r_{23} & \cdots\cdots & r_{2L} \\ r_{L1} & r_{L2} & r_{L3} & \cdots\cdots & 1 \end{pmatrix}$
Inter-cluster correlation matrix.
7