数量化に依る分類の問題

——イネ属を素材として——

(数量化理論とその応用例 (VII))

高　倉　節　子

Some Statistical Methods of Classification by the Theory of Quantification

Setsuko TAKAKURA

In the present paper, some methods of quantification (developed by Hayashi), in the case where many individuals are to be classified into several groups, are described. And then, by the application of these methods, we set forth the consteration of interspecific variation on the genus Oryza. We obtained desirable results from the methodological point of view.

Institute of Statistical Mathematics

§1. はじめに

ここに述べる事は，或集団に於ける“個体”を分類するに際し適用される数量化理論に関し，既に発表された種々の理論及びその展開について説明し，且つそれ等の理論を中心として，野生イネ Oryza 属 16 種に属する 34 系統について，主として形態的特性を用いて行った分類を呈示し，依てこの方法の用途を紹介すると共にイネの分類に関し，一つの試法を呈しようというものである．

数量化を行う場合には，元来，対象とする個体について，目的に対する客観的評価——これを外的規準と名づける——が何等かの意味で与えられていなくてはならない[1]． この方法は目的とする事象を，それに関連する要因からうまく描き出す為の道具であるからである．外的規準は一見与えられていない場合もある．ここで行う分類の様に唯単に要因に於ける反応の模様だけが与えられ，これに基づき似た反応のもの同志を近くに位置づけようという様な場合である．こういう場合にも先づ我々は目的に対し立場を定めてその目的に適う基準を作り，而して後，その規準に適合する方法を見出して行くのが至当であろう．外的規準なくしては方法の妥当性を十分に検討することは出来ないと思われる．所で，ここで素材としてとりあげたイネ属の分類に於てはあらかじめ外的規準をきめるという事は出来なかったので，イネの種・系統の弁別ということを基準にして分類を考えることにした．

さて，ここで旧来行われて来た生物の分類について見て行こう．生物の分類[2]は，初期に於ては或基準となる標本を定めそれと比較して分類を行う“タイプ分類学”が主体となり，これに依り種を他の種から区別しこれに命名し，更には種間の類似等から更に高い分類群にまとめて行くものであった．近代に到って単にその形態的特性を観察するだけではなく，古生物学，比較発生学，細胞遺伝学，血清学等の知見を加えて，系統の概念に基づきその類縁関係を示そうとする“自然分類”の方法に発展し来たり，更に現今では同一種は固定した型（単型）であるとしていた旧時代に代って，

---

1)　林知己夫：　数量化と予測に関する根本概念，彙報7巻1号

2)　加崎英男，岡田豊日：　分類学と系統学：　進化と生命の起源，第3章

種内の変異を分析し種とは多型的なものであるとし，いわゆる「種生態学」的立場から生態型の集団としての種を単位と考え，分類を行うようになった．ここに生物分類の単位となっている“種”なる概念も分類学の発展と共に変化し，画一的な定義は与えられていない．タイプ分類学に於ては“遺伝すべき形態的な特徴により他の種と区別される集団”を以て種と考えていた（これを形態種という）が，自然分類に到って，“生殖の条件，形態と生理，生態等の知見の綜合に依て定められるべき集団”（自然種という）を種としているようである．何れにせよ明確な定義ではない様に思え，又全般的な形質を考慮しても果してどこ迄を同一種と考え何処からを区別される集団と考えるかは曖昧なものであって，分類に於てはこの境界のひき方といい，要因の選び方といい，分類学者の主観の入る余地が多く絶対的な規準をきめる事は困難な様である．

ここで我々が呈示する分類は，自然分類の確立に貢献する事を目的とし，認識し得る形質に依って変異を正確に描写しようとしたものである．即ち，旧来の“種”の概念に把はれず，系統──(現地に野生して一つの集団をなしているものの代表)──を単位とし，測定し得た 42 形質を等価に──今の所，形質の生物学的な重みづけには一定の規準はない──綜合して，系統の相対的位置づけをしてみようという事である．この際，前に述べた意味に於ける外的規準は与えられて居らず，又それをはっきり規定する事も出来ないという重大な点を繰り返しておこう．“類縁関係を示す分類”というものは余りにも多面的なものであり，漠としたものであって，この為のはっきりした規準は把握出来ないものである．即ち何と何とが近いとか遠いとかいった時に，その系統的な意味での近さとはどういう場合であるかという事は我々の具体的・即物的・操作的目的がはっきりと手にとれる形で定められない限り規定する事が出来ないものであり，現在の場合，その目的が定められない場合であるから外的規準をきめることは困難である．それでここで行った分類は，静的なものであり，生物変異の鳥瞰図を立場を定めて描いてみようという事でしかない．この立場のとり方は，いろいろ考えられるが，我々としては統計数理的な考え方から現在において可能なる限り妥当あらしめようとつとめてみた．この様にして，複雑なデーターの様相を解り易く描いてみせる一つの試法であって，この描かれたものを呈示する事に依て，当該現象に関する一つの知識（評価づけられた行動規範とはならない）として，斯学の進歩の足がかりともなろうと思うのである．

§2. 素材とした資料

ここで取上げた資料は，国立遺伝学研究所から供せられたものであって，分類の対象はイネ属 (Oryza) 20 種余りの中 16 種 34 系統である．1つの種には1系統しか測定していないものもあり，又数系統について測定値を得ているものもある．対象とした系統は次頁の通りである．

以下に行う分析に於ては常に系統を単位として，取扱う事としている．系統を分類するに当って用いた要因は次の様な 42 の形質である．これ等の形質はそのすべてを同じ個体について測定したわけではない．然し，少くとも質的特性は，同一系統であるならば，どの個体で測っても同じカテゴリーに属するものであるといわれて居り（遺伝学研究所森島啓子氏による）又量的特性についてはこのデーターでは一応個体に依る差異はとり上げず，5個体以上の測定値の平均を系統の代表値としたものである．

この事は不十分であって，もっと正確な結果に達するためには同一系統内個体間の測定値の変動がどれ位起り，また環境に依てどれ位異なるか等を充分に把えておく必要があり，又分析に用いる数値の安定性の為にも，基礎的データーとして，多数の個体についての測定値から分析を始めるべきであろう．尚，ここで単位にとった系統は，同一種の中に多数の系統があるものもあり，又採集困難なる場所に於ける種では小数系統しかとられていない．この様な事も統計的処理を行う場合には障害となる事であり，今後検討を要する事となろう．然しとも角，ここでは一応供せられた測定値を各系統の形質の測定値として出発した．従って，このデーターに基づいて行った以下の分析に

測定値を得た系統 No.		所属する種	No.
1	108		
2	414		
3	521	O. sativa L.	1
4	563		
5	647		
6	W 0106	O. f. spontanea Roschev.	2
7	W 0122		
8	W 0032		
9	W 0126	O. perennis Moench*	3
10	W 0149		
11	W 0040		
12	W 0492	O. glaberrima Steud.	4
13	W 0493		
14	W 0009		
15	W 0042	O. breviligulata A. Cheval et Roehr.	5
16	W 0049		
17	W 0019		
18	W 0020	O. latifolia Desv.	7
19	W 0047		
20	W 0021	O. malabarensis	12
21	W 0016	O. minuta Presl.	9
22	W 0045		
23	W 0002		
24	W 0006	O. officinalis Wall.	10
25	W 0012		
26	W 0008	O. australiensis Domin	6
27	W 0015	O. eichingeri Peter.	11
28	W 0043		
29	W 0017	O. alta Swallen	8
30	W 0048		
31	W 0023	O. brachyantha A. Cheval et Roehr.	15
32	W 0003	O. granulata Nees	13
33	W 0001	O. ridleyi Hook.	14
34	W 0510	O. subulata Nees	16

* この種には
O. barthii A. Cheval
O. longistaminata A. Cheval
O. cubensis Ekman
を含む

形質 No.	定量的形質	単位	形質 No.	定性的形質	カテゴリー 0	1	2	3
1	籾の長さ	mm	21	葉舌の形		長	中	短
2	〃 幅	mm	22	〃 毛		無		有
3	〃 長×幅		23	〃 硬さ		膜質		硬
4	〃 長/幅		24	葉耳の形	無	小	中	大
5	稃毛の長さ	mm	25	第3葉の幅		細	中	広
6	護穎の長さ	mm	26	第3葉の角度		小	中	大
7	葉舌の長さ	cm	27	分けつの角度		直	中	広
8	気孔細胞の長さ	μ	28	草丈		低	中	高
9	籾長/葯長		29	節間の節		無		有
10	花粉直径	μ	30	穂の構造		1本		2本以上
11	芒の長さ	cm	31	枝梗の太さ		細	中	太
12	脱粒性指数		32	〃 毛		無	中	有
13	アルカリテスト指数		33	穂の形		狭	中	広
14	一次枝梗数		34	ペデイセルの形				
15	着粒密度		35	護穎の形		舟形		針状
16	一穂粒数		36	〃 毛		無		有
17	穂長	cm	37	内穎の伸長		普通		伸
18	100粒重	gm	38	籾の毛の分布	無毛	全体		脈に沿って
19	発芽率		39	芒の太さ		細	中	太
20	発芽日数		40	地下茎		無	中	有
			41	一年生, 多年生		一年生		多年生
			42	幼植物の生長速度		遅	中	早

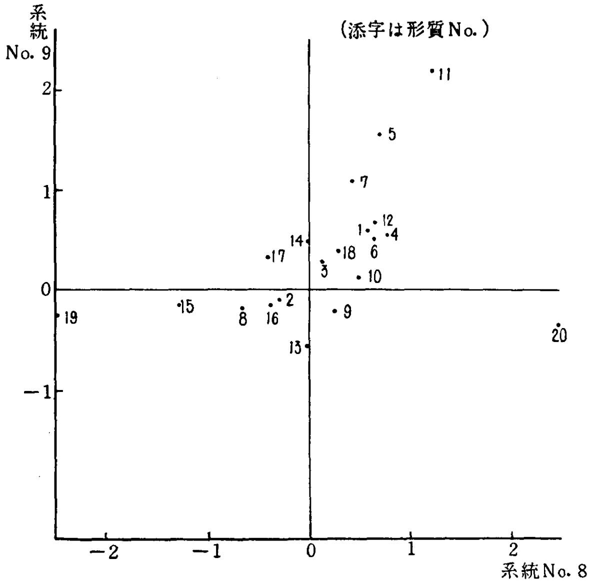
依る結果だけから，どの系統とどの系統は同一種と見做すべきであるか否か，という断定的な結論を下す事は困難であるし，又新たな系統を持って来てここで取上げた形質を測定した結果，“これは何々の種である”と判断を下す様な予測に役立てる程，この結果の数値そのものに安定性はないと思われる．つまりここでは，研究の方法と一応の結果とを呈示するに留まるが，今後十分なるデーターから出発するならば，可成り安定したものとして，変異の様相を示すことが出来よう．

形質は主として形態的特性であるが 13，19，20，42 等の様な生理的特性も加はっている．元来，自然分類を行のうに当っては多数の形質を考察して行うべきであるが，ここでは一応無数の形質の中から何等かの意味で（主として測定され易いために）選ばれた形質を用いた．この中 1～20 は数量を以て測定される特性であるが，特性に依て測定の単位が異なるので，以下の分析に於ては各特性別に平均 0，分散 1 に変換した値を用いる事とした．これに反し，21～42 は定性的な特性であってその右欄のカテゴリーで記述されている．これ等は便宜上最上欄のコードで表わす事とし以下の分析にはこのコードを用いている．

§3. 解析の始めに

先づ最初に全般の様相を見透す為に，系統間，及び形質間の関係を図示してみた．

先づ系統相互間の関係を見るに，任意の2本の系統をとり上げ（すべての組合せをとった）これを X 軸，Y 軸に対応させ，20 個の量的形質夫々に関して2つの系統 X, Y の測定値を図示した．



その一例を示せば右図の通りである．これに依ると，同一種内の系統はやはり全般的に可成り高い相関を示すことがわかる．但し，この関係は普通の意味での相関とは少し違う．即ち一般には，着目する或特性（標識）について，サンプル測定値をユニットと考えた場合の相関であるが，ここでは逆にユニットとしてとるものは形質であり，その標識として系統名が与えられている．形質は gm, cm 等いろいろの単位で測定されるものであるが，これを同じ尺度としてみる必要があったので，一応各々平均 0，分散 1，に変換することとし（これについての是非はここではとり上げない），これ等の尺度に依るすべての形質を母集団と考え，これから抽出された形質であると考えるのである．次に同様の事を定性的特性について行ってみた．この場合，測定は数量ではなくコードで表わされているが，数量の場合と同様に2つの系統をとり上げ，定性的形質をユニットとし，2系統の反応コードに依る頻度表を作った．ここで考慮すべき事は，コードの付け替えに依て分布の様相が変るという事である．依て，これから何等かの関連の尺度を作る場合には注意しなくてはならない．次に見方を変えて形質間の相関を描いてみる．定量的形質，定性的形質のすべての中から2つづつのあらゆる組合せをとり上げ，これを2次元の軸にとりすべての系統をプロットしてみた．これによると，定性的形質の2つをとり上げたものの中には各々の桝目の中に同一種に属する系統が入る事が概して多い様であり，定性的特性は系統の分類に可成り力がある様に見うけられる．又形質の間で可成り強い関連のあるものも見うけられた．例えば｛葉舌の形と硬さ｝，｛葉舌の形と長さ｝，｛籾の長×巾と籾長/籾長｝，｛葉の角度とペディセル

の形}，{幼植物の生長速度とペディセルの形}，{穂長と一穂粒数}，{着粒密度と一穂粒数} 等である．これ等は生物学的に当然であるものもあろうし，又何の関係も考えられぬものもあるかも知れない．もし生物学的にはっきりした裏づけがあるならば，今後分類の要因とする特性を選ぶ時に考えるべきであろう．

これ等は全体の見透しをよくする為のものであって以下に述べる解析の足がかりとなるものである．

§4. 数量化理論の説明とその適用

I. e_{ij} 型数量化[3)]

これはソシオメトリーに於けるグルーピングの問題から発生したもので或集団に於て，2つづつの組合せをとり上げ，何等かの尺度でその相互の親近度が測定出来た時に，その尺度を用いこの集団全体を総括した上で，親近度の強いものは近くに薄いものは遠くになるように各個の相対的位置づけを行うものである．即ち

e_{ij}： i なる個の j なる個への親近度，e_{ij} の大なるほど親近度が高いとしておく．

$$i \neq j,\ i, j;\ 1, \cdots, N \quad (N\text{: 個体の総数})$$

x_i： i なる個に与える数値

としたとき，

$$Q = -\sum_{i=1}^{N}\sum_{j=1}^{N} e_{ij}(x_i - x_j)^2$$

なるものを尺度とし，x の分散一定の下に，Q を最大ならしめる様な x を求めるのである．つまり e_{ij} が大なれば大なる程 x_i と x_j との距離は小となる様に x の座標となる数量を与えようという事である．

これを実際に計算するには，

$$\mathrm{A}\boldsymbol{x} = \lambda\boldsymbol{x} \quad \boldsymbol{x}' = (x_1, x_2, \cdots, x_N)$$

なる固有方程式を解けばよい．この時 A の i 行 j 列の要素を A_{ij}，i 行 i 列の要素を A_{ii} とすれば

$$A_{ij} = e_{ij} + e_{ji} \qquad e_{ij} = e_{ji} \text{ の時は } A_{ij} = e_{ij}$$
$$A_{ii} = -\sum_{j \neq i} A_{ij} \quad \text{である．}$$

尚この時の最大固有値 ${}^1\lambda$ に対応する固有ベクトル ${}^1\boldsymbol{x}$ のエレメント ${}^1x_i$ は個体の距離を1次元上に考えた場合，最適の個体の位置を表わす座標であり，${}^1\boldsymbol{x}$ に直交せる第2根 ${}^2\lambda$ に対応する固有ベクトル ${}^2\boldsymbol{x}$ を求めれば ${}^1x_i, {}^2x_i$ は，2次元上の個間の距離を考えた時に最適になる様な個体の位置を示す直交軸の座標となる[4)]．

この場合計算上の注意としては固有値は一般に正とは限らず（$A_{ij}(i \neq j)$ がすべて $\geqq 0$ なる時は λ はすべて $\leqq 0$ である），然も rank は少くとも1つ落ちているので，固有値の大なる順に求めて行く時には $\lambda = 0$ なるものに出会う．従って少くとも欲する根迄は $\lambda \geqq 0$ となっている様に，始めの要素を変換しておく方が便利である[5)]．即ち $c<0$ なる任意の常数 c をとり

$$A_{ij}' = A_{ij} + c$$
$$A_{ii}' = -\sum_{j \neq i} A_{ij}' = -\sum_{j \neq i} A_{ij} - (N-1)c$$

としてから出発する．この様な変換を行っても求めるベクトルには関係ない．この時 $|c|$ を余りに

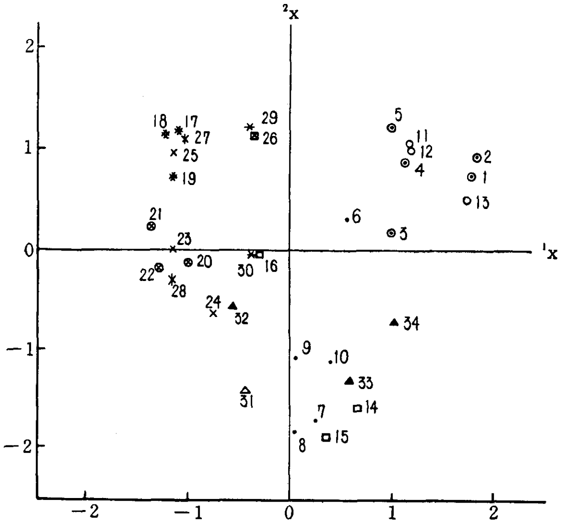
3) C. Hayashi: On the Prediction of Phenomena from Qualitative Data and the Quantification of Qualitative Data from the Mathematico-statistical point of view, Ann. I.S.M. Vol. III, No. 2 (1952)
林知己夫，他： 態度数量化の一方法 (II)，彙報6巻1号

4) 林知己夫： 数量化理論とその応用例 (VI)，彙報9巻1号

5) 林知己夫，他： 態度数量化の一方法 (II)，彙報6巻1号

大きくとると収斂が遅くなる．

さて，ここで e_{ij} として次の様な尺度をとる事とした．



(i)　相関係数を e_{ij} とする．

系統相互間の近さの尺度として，2つの系統間の量的特性に関する相関係数をとる事とした（この場合，量は当然各特性毎に平均 0，分散 1 に変換された値を用いている）．これに依り最大根と第2根とを求め，これに見合う2つの固有ベクトル $^1\boldsymbol{x}, {}^2\boldsymbol{x}$ を求め，これから2次元の位置づけを行ってみた．これは左図の如くである．

(ii)　カテゴリーの一致度を e_{ij} とする[6]．

前述の相関係数は質的特性には適用出来ない．カテゴリーがコードで表現されているのであるから数量として扱えず，相関係数を計算する事は出来ない．依て定性的形質に関して，親近度を表わすものとして，次の尺度を用いてみた．即ち解析の始めに作った2つの系統に依る，すべての定性的形質に関する相関表を用いて計算を始めるのであるが，先づ次の様に定義をする．即ち

$f_{ii}(l,m)$：　系統 l, m が共に i-category に反応した形質の総数，

$f_{i\cdot}(l)$：　系統 l が i-category に反応した形質の数，

n：　形質の総数，

とする．今，l,m なる2つの系統について，夫々の系統における形質の分布が与えられた時に，これ等2系統が確率的に独立であるならば2系統が共に i-category に反応する確率は $f_{i\cdot}(l)f_{i\cdot}(m)/n^2$ であり，ここにおちる形質の数 $f_{ii}(l,m)$ は二項分布をなし，その平均は $f_{i\cdot}(l)f_{i\cdot}(m)/n$，分散は，

$n\dfrac{f_{i\cdot}(l)f_{i\cdot}(m)}{n^2}\left(1-\dfrac{f_{i\cdot}(l)f_{i\cdot}(m)}{n^2}\right)$ である．ここで

$$e_{ii}(l,m)=\frac{\left(f_{ii}(l,m)-\dfrac{f_{i\cdot}(l)f_{i\cdot}(m)}{n}\right)}{\sqrt{n\cdot\dfrac{f_{i\cdot}(l)f_{i\cdot}(m)}{n^2}\left(1-\dfrac{f_{i\cdot}(l)f_{i\cdot}(m)}{n^2}\right)}}$$

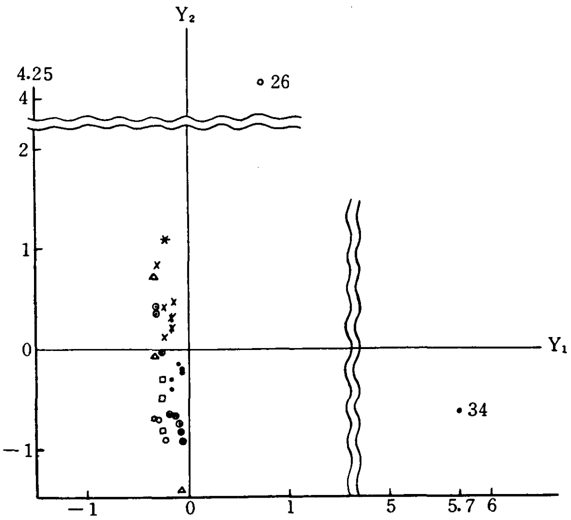
なる尺度を定義すれば，これは系統 l, m が，夫々独立に各形質のカテゴリーに反応すると仮定した時，同じカテゴリーに反応した頻度が，その平均的に起ると考えられる頻度に比べ，どれ程距たっているか（分散一定とした時）の尺度であり，これが0なる時は l, m は独立であり，大なる程同じカテゴリーに反応し易い．つまり関連があるという事である．これを総てのカテゴリーについて綜合し，即ち

$$e(l,m)=\sum_i e_{ii}(l,m)$$

と定義し，この $e(l,m)$ を以て l,m の一致度，つまり結びつきの尺度とする．

これを質的特性に適用して，前述の如く，固有ベクトルを求め，2次元に示せば次図の如くである．これによると，系統 No. 26 と No. 34 の系統が異質であり，これが余りにも他と異った為，これを取出だす事だけで，他の系統は殆ど全部同位置となっている．従って，質的特性から見る場合 No. 26，No. 34 は特異なものである事が解ったので，これは他の 32 系統とは別個のものとし

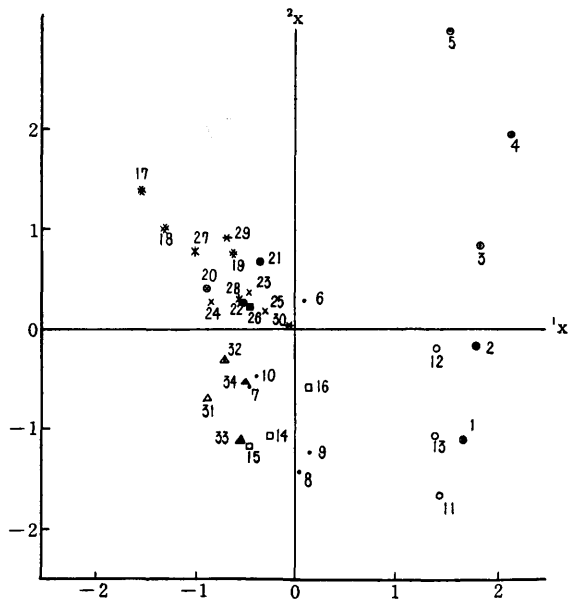
6)　5) と同じ．



て扱った方がよい様に思える（尚，この2系統を別にして残りの 32 系統について，この方法を再び適用する事は後述の理由に依り省略した).

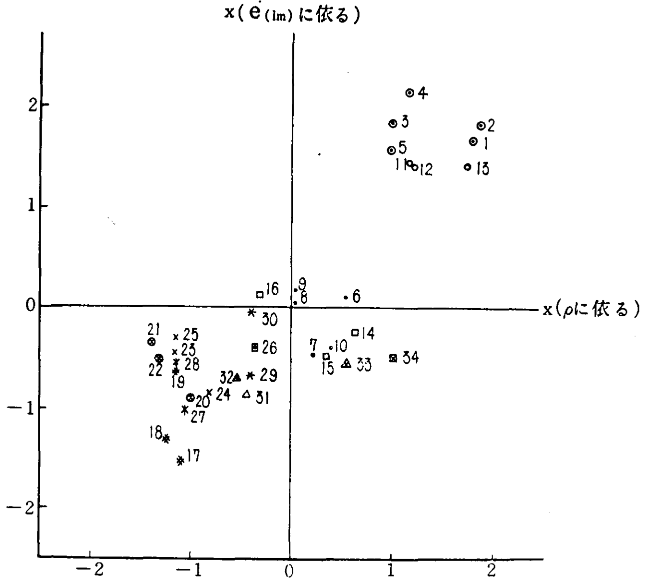
次にこの方法を量的特性についても行った．これには，先づ量を幾つかのカテゴリーに分ける為に2系統間の相関図，量的特性の分布及び分散等から区切りを次の如くに決めた．即ち

（−1.5 未満)，(−1.5〜−0.5)，(−0.5〜0.5)，(0.5〜1.5)，(1.5 以上）の5カテゴリーに分け質的特性と同様の事をした．この結果は次図の通りである．



量的特性について，前の相関係数を用いたものと $e(l,m)$ を用いた場合との X の関係を2次元に図示してみると可成り高い相関がみられ，何れの方法によろうとも，この場合は結果として余り違わない事が解る．

$e(l,m)$ なる尺度は，質，量何れの特性にも用い得る尺度で，これに依り，質，量全形質を綜合的に用いて系統の位置づけを行はうとしたが，質的特性に適用する場合には，先にも触れた様にコ



ードのつけ変えについて問題がある．即ち，カテゴリーをコードで表わす際に，例えば，“薄い”“普通”“厚い”という事を 0, 1, 2 とするが，この際どちらを0とするかは全く任意である．然も方向づけを任意にきめる事に依て，周辺分布は異った様相を呈する．つまりコードのつけ変えに依て，$\sum f_{ii}(l,m)$ なる和は変らないが，前記の定義による $e(l,m)$ なる値は異った値となる．

従ってこの尺度をとる場合には，コードの方向づけを逆にした場合をとり上げ，その総ゆる組合せによる $e(l,m)$ の平均値をこの尺度としてとるべきであろう．然し，それは余りに煩雑であって計算に手間どる事となる．依てこの方法による分析はこれに留め，この考え方を更に発展させた方法の適用に移ろう．

II. 質的又は量的関係よりする数量化による分類——多次元的コンステレイションの描写

始めに方法の概要を述べ，次に適用に移る．

[方法の概要]

(i) 要因が質的な場合

——反応パタンによる方法[7]——

或集団に於ける個体の，幾つかの要因に対する反応パタンが示された時(左図)，同じ様な反応を示したものは近くに，異った反応のものは遠くになるように，即ち右側の図に示す様に，要因と個とを直線的に関係づけ，相関係数を max. とする様に，それ等に数量を与えようという事である．即ち，

個＼要因	1	2	3	4	5
1	∨				∨
2	∨		∨		∨
3			∨		∨
4			∨	∨	∨
5			∨	∨	
6		∨		∨	

個＼要因	1	5	3	4	2
1	∨	∨			
2	∨	∨	∨		
3		∨	∨		
4		∨	∨	∨	
5			∨	∨	
6				∨	∨

$$\left.\begin{array}{l}\text{要因に与える数量を } x\\ \text{個に与える数量を } y\end{array}\right\}\text{としたとき，}$$

相関係数 ρ_{xy} を max. とする様な x, y を求める事である．この事は，$(1-\rho_{xy}^2)$ を min. にする事であり，x, y の何れか一方，例えば x に関し，全体の分散に対する内分散を min. にする，即ち個体を夫々層として見た時の相関比 η を max. とする要因の数量 x を求める事であり，且つ，x, y を線型の関係にするのであるから y の値は各個 (i) に於ける x の平均値を与えればよい事になる．相関比を max. にする x を求める方法を数式で示せば次の通りである．

7) 林知己夫：数量化理論とその応用例 (II)，彙報4巻2号

N なる個体の集団を R 個の要因に就て反応を調べ，i なる個の j なる要因への反応を $\delta_i(j)$ にて表わし，

$$\left.\begin{aligned}\delta_i(j)=1:&\quad \text{反応を示したとき,}\\ 0:&\quad \text{反応を示さなかったとき,}\end{aligned}\right\}\text{と定義する.}$$

$$f_{\cdot j}=\sum_i \delta_i(j):\quad j\ \text{要因が反応を示された総数}$$

$$f_{i\cdot}=\sum_j \delta_i(j):\quad i\ \text{なる人が反応を示した総数}$$

$$T=\sum_j f_{\cdot j}=\sum_i f_{i\cdot}\quad \text{とする.}$$

要因 j に与える数値を x_j とし，個体 i には

$$\bar{x}_i=\frac{1}{f_{i\cdot}}\sum_j x_j\delta_i(j),$$

即ち i に於ける x の平均値を与えた時，個の間の variance，即ち x の between の variance は

$$\sigma_b{}^2=\frac{1}{T}\sum_i\frac{1}{f_{i\cdot}}\Big(\sum_j x_j\delta_i(j)\Big)^2-\Big(\frac{1}{T}\sum_j x_jf_{\cdot j}\Big)^2$$

matrix 表示をすれば

$$T\sigma_b{}^2=\boldsymbol{x}'H\boldsymbol{x},\quad \text{但し}\ \boldsymbol{x}'=(x_1,\cdots x_j,\cdots,x_R)$$

H の j 行 k 列の要素を H_{jk} とすれば

$$H_{jk}=\sum_i\frac{\delta_i(j)\delta_i(k)}{f_{i\cdot}}-\frac{f_{\cdot j}f_{\cdot k}}{T}$$

x の全体の分散は

$$\sigma^2=\frac{1}{T}\sum_j x_j^2f_{\cdot j}-\Big(\frac{1}{T}\sum_j x_jf_{\cdot j}\Big)^2$$

$$T\sigma^2=\boldsymbol{x}'F\boldsymbol{x}$$

F の j 行 k 列の要素を F_{jk} とすれば

$$F_{jk}=-\frac{f_{\cdot j}f_{\cdot k}}{T}$$

$$F_{jj}=f_{\cdot j}-\frac{f_{\cdot j}f_{\cdot j}}{T}$$

そこで $\eta^2=\dfrac{\sigma_b{}^2}{\sigma^2}=\dfrac{\boldsymbol{x}'H\boldsymbol{x}}{\boldsymbol{x}'F\boldsymbol{x}}\to\max.$ なる様な $\boldsymbol{x}$ を求めると，これは $\dfrac{\partial\eta^2}{\partial\boldsymbol{x}}=0$ より

$$H\boldsymbol{x}=\eta^2F\boldsymbol{x}$$

なる固有方程式を満たす任意解を求めればよい事になる．ここに x の全体の平均値を任意におく事が出来るのでそれを 0 とする．

尚，$H,F,$ なる行列は夫々

$$H=\frac{1}{R}GG'-\frac{1}{T}\boldsymbol{ff}'$$

$$F=D-\frac{1}{T}\boldsymbol{ff}'$$

なる型に表す事が出来る．

ここに G は R 行 N 列の matrix で，その j 行 i 列の要素は $G_{ji}=\delta_i(j)$，D は対角行列で，要素は $D_{kk}=f_{\cdot k}$，$\boldsymbol{f}$ はベクトルで $(f_{\cdot 1}\cdots f_{\cdot j}\cdots f_{\cdot R})'$ を表わす．

そこで全体の平均を 0 とおく事は

$$\boldsymbol{f}'\boldsymbol{x}=0$$

とすることである．

こうすれば $\boldsymbol{Fx}=\boldsymbol{Dx}$ となり，$\boldsymbol{Hx}=\eta^2\boldsymbol{Fx}$ に於て，$\boldsymbol{f'x}=0$ を満足する解は $\boldsymbol{Hx}=\eta^2\boldsymbol{Dx}$ から得られることとなる．何故ならば，右辺の各行をすべて加え合せれば

$$\sum_k D_{kk}x_k=\sum_k f_{\cdot k}x_k$$

となり今0とおいた $\boldsymbol{f'x}$ と等しくなり，又，左辺の各行をすべて加え合せれば

$$\sum_j \sum_k H_{jk}x_k=\sum_k (\sum_j H_{jk})x_k$$

で $\sum_j H_{jk}=0$ であるから，常に0となる．依て

$$\boldsymbol{Hx}=\eta^2\boldsymbol{Dx}$$

は $\boldsymbol{f'x}=0$ を満足する $\boldsymbol{Hx}=\eta^2\boldsymbol{Fx}$ の解となっている．

（尚 $\boldsymbol{Hx}=\dfrac{1}{R}GG'\boldsymbol{x}$ となるが $GG'\boldsymbol{x}=R\eta^2D\boldsymbol{x}$ の解は $\boldsymbol{f'x}=0$ を満足しない）

$$\boldsymbol{Hx}=\eta^2\boldsymbol{Dx}$$

から解く事は D が対角行列であるから計算が簡単となり便利である．これから x を求め，y は

$$y_i=\frac{1}{f_{i\cdot}}\sum_j x_j\delta_i(j)$$

に依て定められる．

ここで，この場合の固有値，並びに固有ベクトルの意味について触れよう[8]．固有値の第1根は $\bar{x}_i$ の値を1次元に投影した時の分離度となる相関比 η_1 であり，固有ベクトルはその軸を表わし，第2根迄求めるという事は2次元における夫々の直交軸に投影した場合の相関比 ${}^1\eta, {}^2\eta$ の積，即ち ${}^1\eta, {}^2\eta$ で囲まれた面積を max. にする，いいかえれば2次元に於ける分離度を最大にする為の直交軸を求める事である．後述の様に数量化の特別な場合は因子分析法に於ける主因子解と式の型は全く同じになるが，数量化の場合は，軸の解釈がはっきりしている．即ち一般に p 根迄固有値及び固有ベクトルを求めるという事は，p 次元に於ける分離度（各軸上の分離度の積を以て p 次元の分離度と考えた場合）が最大なる様な直交軸を選ぶという事であり，これに反して因子分析の場合はとり出した因子軸に意味がないのである．

(ii) 要因が量的な場合（その 1)[9]

今度は要因が定量的なる場合を考える．即ち i なる個体の j なる要因に於ける測定値（数量）を u_{ij} とする．この値を要因別に1次変換して新たな数量 x_{ij} を与えたとした時，この x_{ij} に対して前と同様に η^2 max. となる様な一次変換のパラメーターを決めようという事である（ここに個体の数は N，要因は R 個とする)．即ち

$$x_{ij}=a_ju_{ij}+b_j \quad \text{とした時}$$

$\eta^2=\dfrac{\sigma_b{}^2}{\sigma^2}\to$max. となる a_j, b_j を決めることである．ここで

$$\frac{\sigma_b{}^2}{\sigma^2}\to\text{max.} \quad \text{とすれば} \quad \bar{x}_j=\bar{x}$$

となる．即ち各要因毎の平均値は相等しく全体の平均と一致する．前と同様 $\bar{x}=0$ とおけるから

$$b_j=-\frac{1}{N}a_j\sum_i u_{ij}$$

として求まる．依て

$$x_{ij}=a_j\left(u_{ij}-\frac{1}{N}\sum_i u_{ij}\right)$$

とおき，この a_j をきめる問題となる．これは次の様にする．

---

8) 林知己夫：数量化理論と応用例 (VII)，彙報9巻1号

9) 林知己夫：数量化理論とその応用例 (V)，彙報8巻2号

$$\bar{x}=\frac{1}{R}\sum_{j=1}^{R}a_j\left(u_{ij}-\frac{1}{N}\sum_i u_{ij}\right)$$

$$R^2N\sigma_b{}^2=\boldsymbol{a}'H\boldsymbol{a}$$

ここに

$$\boldsymbol{a}'=(a_1, a_2, \cdots, a_j, \cdots, a_R)$$

H の要素を H_{jk} とすれば

$$H_{jk}=\sum_i u_{ij}u_{ik}-\frac{1}{N}(\sum_i u_{ij})(\sum_i u_{ik})$$

$$RN\sigma^2=\boldsymbol{a}'F\boldsymbol{a}$$

F の要素を F_{jk} とすれば

$$F_{jk}=0$$

$$F_{jj}=\sum_i u_{ij}{}^2-\frac{1}{N}(\sum_i u_{ij})^2$$

$$\eta^2=\frac{\boldsymbol{a}'H\boldsymbol{a}}{\boldsymbol{a}'F\boldsymbol{a}}$$ を max. とする解は

$$H\boldsymbol{a}=R\eta^2F\boldsymbol{a}$$ を解けばよい．

ここで u_{ij} を各 j 項目毎に平均 0，分散 1 に変換しておくならば，即ち

$$u_{ij}'=\frac{u_{ij}-\bar{u}_j}{\sigma_{u_j}}$$ とするならば

$$x_{ij}=a_j\sigma_{u_j}u_{ij}'$$ となり

$$a_{(0)j}=a_j\sigma_{u_j}$$ とおけば，

同様の手段で

$$H_{(0)}\boldsymbol{a}_{(0)}=NR\eta^2\boldsymbol{a}_{(0)}$$

を解けばよい事になる．ここに

$$H_{(0)jk}=\sum_i u_{ij}'u_{ik}'$$

となる．これより求めた $a_{(0)}$ を u_{ij}' と一次結合する事に依て x_{ij} がきまる．尚，この式は因子分析法に於て，コミュナリティを 1 とした時の主因子解と全く同じ式となっている．

因みに次の事をつけ加えておこう．u_{ij} が 1，0 の数値のみをとるものとする．即ち (i) の反応パタンに依る場合の如く $\delta_i(j)$ であったとする．この時

H の要素は

$$H_{ik}=\sum_i \delta_i(j)\delta_i(k)-\frac{1}{N}f_{\cdot j}f_{\cdot k}$$

F の要素は

$$F_{jj}=f_{\cdot j}-\frac{1}{N}f_{\cdot j}f_{\cdot j},\quad F_{jk}=0$$

となる．即ち，反応パタンの場合に，個人の反応の総数 $f_{i\cdot}$ がすべて相等しかったならば，H はこれと同じになる．又 F は全体の分散を計算する場合の x にかかる重みとなるものであるが，これは，反応パタンの時は j カテゴリーに反応する頻度となって居り，この場合には（(ii) に即応して考えるならば）反応パタンの分散となっている事が解る．即ち，前記の反応パタンの場合に於て，排反する行為も同時にとり上げて結びつきを見ようとするならば，――即ち或項目に対して“反応を示した”，“反応を示さなかった”という 2 つの行為を全く等価にとり扱って両方のカテゴリーに対してのパタンから数量化しようとするならば――この量的要因の重みづけと同じ事になるのである（この事から (i) の方法と因子分析法との関連が窺えよう）．(i) の方法は積極的に“反応を示した”もののみを浮彫りにした上での結びつきであり，我々がとり上げる問題如何に依て，例えば社会調査の問題等であれば質問の作り方に依て，何れか適切なる方法をとり上げるべきであろう．

尚，後述する様にイネの分類の問題で (i) の方法を用いるのであるが，ここでは項目の中を排反

せる数個のカテゴリーに分け，その反応パタンから結びつきを見るという方法をとっている．この場合には H の matrix は (ii) の場合と同様になり，全体の分散は (i) の様に頻度に依る重みづけとなっている．然しこの場合は，項目内は排反事象となっているのであるから，そのカテゴリーの“起り易さ”に対して“起り難さ”という面も別のカテゴリーから加味されてくるので (i) と同じ手法を使った方が適切と考えられ，然も，これは量的にとり扱った方法 (ii) と同じ意味を含むものであるから，(iv) で述べる様にこれに量的特性をも同時に並べ，この (ii) 又は次の (iii) の方法とを併せて用いても不自然はないと考えられよう．

(iii) 要因が量的な場合（その 2）

上の方法から更に発展させ，定量的な要因の場合，初めの測定量を u_{ij} としたとき，

$$x_{ij}=a_ju_{ij}^2+b_ju_{ij}+c_j$$

なる2次変換を行って，新たな x_{ij} を与え，この値に依る η^2 が max. になる様な2次式のパラメーターを求める事とする．前記 (ii) の方法では各項目毎の重みを決める事であるので，項目間のきき方を調整するだけである．然し各項目に於ける測定量が個体を分離するという事に対し，一次の関係にあるとは限らないので——一般にはいろいろ複雑な型をしていようが——これを稍々補正する意味に於て2次式を想定してみたのである．例えば 1 cm と 3 cm とは大差なしと判断される事であり，3 cm と 5 cm とは非常に違う事であるという様な要因もあろう事を想定したのである．

この場合も前と同様，各要因毎の平均は総て等しくなり，又これを0とおいても差支えない．即ち $\bar{x}=0$ とおけるのであるから

$$c_j=-\left(a_j\frac{1}{N}\sum_i u_{ij}^2+b_j\frac{1}{N}\sum_i u_{ij}\right)$$

として求まる．依て始めの2次式は

$$x_{ij}=a_j\left(u_{ij}^2-\frac{1}{N}\sum_i u_{ij}^2\right)+b_j\left(u_{ij}-\frac{1}{N}\sum_i u_{ij}\right)$$

となる．これから a_j, b_j を次の様にして求める．

$$\bar{x}_i=\frac{1}{R}\sum_j a_j\left(u_{ij}^2-\frac{1}{N}\sum_i u_{ij}\right)-b_j\left(u_{ij}-\frac{1}{N}\sum_i u_{ij}\right)$$

$$NR^2\sigma_b^2=\begin{pmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{pmatrix}'\Big(H\Big)\begin{pmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{pmatrix}$$

ここに

$$\boldsymbol{a}=(a_1, a_2, \cdots, a_j, \cdots, a_R)'$$
$$\boldsymbol{b}=(b_1, b_2, \cdots, b_j, \cdots, b_R)'$$

H は $2R\times 2R$ の行列であり，その要素は

$$\Big(H\Big)=\begin{pmatrix}H^{(11)} & H^{(12)}\\ H^{(21)} & H^{(22)}\end{pmatrix}$$

として各部分は夫々 $R\times R$ の行列でその j 行 k 列は

$$H_{jk}^{(11)}=\left(\sum_i u_{ij}^2u_{ik}^2-\frac{1}{N}\sum_i u_{ij}^2\sum_i u_{ik}^2\right)$$

$$H_{jk}^{(12)}=\left(\sum_i u_{ij}^2u_{ik}-\frac{1}{N}\sum_i u_{ij}^2\sum_i u_{ik}\right)$$

$$H_{jk}^{(21)}=\left(\sum_i u_{ij}u_{ik}^2-\frac{1}{N}\sum_i u_{ij}\sum_i u_{ik}^2\right)$$

$$H_{jk}^{(22)}=\left(\sum_i u_{ij}u_{ik}-\frac{1}{N}\sum_i u_{ij}\sum_i u_{ik}\right)$$

又，

$$NR\sigma^2=\begin{pmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{pmatrix}'\Big(\boldsymbol{F}\Big)\begin{pmatrix}\boldsymbol{a}\\ \boldsymbol{b}\end{pmatrix}$$

同様に部分に分けてそれぞれの要素は

$$F_{jj}^{(11)}=\sum_i u_{ij}^4-\frac{1}{N}(\sum_i u_{ij}^2)^2$$

$$F_{jj}^{(12)}=F_{jj}^{(21)}=\sum_i u_{ij}^3-\frac{1}{N}(\sum_i u_{ij}^2)(\sum_i u_{ij})$$

$$F_{jj}^{(22)}=\sum_i u_{ij}^2-\frac{1}{N}(\sum_i u_{ij})^2$$

$$F_{jk}^{(11)}=F_{jk}^{(12)}=F_{jk}^{(21)}=F_{jk}^{(22)}=0$$

ここで

$$R\eta^2=\frac{\begin{pmatrix}\boldsymbol{a}\\\boldsymbol{b}\end{pmatrix}'\begin{pmatrix}H\end{pmatrix}\begin{pmatrix}\boldsymbol{a}\\\boldsymbol{b}\end{pmatrix}}{\begin{pmatrix}\boldsymbol{a}\\\boldsymbol{b}\end{pmatrix}'\begin{pmatrix}F\end{pmatrix}\begin{pmatrix}\boldsymbol{a}\\\boldsymbol{b}\end{pmatrix}}$$

を max. にする解は

$$(H)\begin{pmatrix}\boldsymbol{a}\\\boldsymbol{b}\end{pmatrix}=R\eta^2(\boldsymbol{F})\begin{pmatrix}\boldsymbol{a}\\\boldsymbol{b}\end{pmatrix}$$

を解けばよい.

これから $\boldsymbol{a}, \boldsymbol{b}$ を求め，新たな x_{ij} が決まる．尚ここで H は，次の様に行列の積に依て，表わされる.

$$H=MM'$$

ここで M は $2R$ 行 N 列の行列であり，その j 行 i 列及び 2_j 行 i 列の要素は夫々

$$M_{ji}=u_{ij}^2-\frac{1}{N}\sum_i u_{ij}^2$$

$$M_{2j\,i}=u_{ij}-\frac{1}{N}\sum_i u_{ij}$$

この事を利用すると，後に述べる計算法上の問題で便利な事がある.

(iv) 要因に質的なものと，量的なものとの両方を含む場合（その 1）

前述の様に (i) の方法も排反せる数個のカテゴリーからなる項目への反応パタンをとり上げた場合はパタンを量とした場合の変形として考えられるから，これと (ii) の方法とを併せて量化を行っても不自然はない．即ち，i なる個が質的要因 j の m カテゴリーに対しては $\delta_i(j_m)$ の反応を示し，量的要因 l に於ける測定量を u_{il} とする（ここで前と同様各要因毎に平均 0，分散 1 に変換しておき，これを u_{il}' とする)．質的要因は R 個，j なる要因に於けるカテゴリー j_m は夫々 j_M 個づつあるとして，カテゴリー総数を $T\,(T=\sum_{j=1}^{R}\sum_{j_m=1}^{j_M} j_m)$，量的要因は S 個であったとする.

尚，j-item m カテゴリーに反応した頻度を (i) と同様に $f_{\cdot j_m}$ で表す ($f_{\cdot j_m}=\sum_i\delta_i(j_m)$．この時 i なる個に与えられる数量は

$$\bar{x}_i=\frac{1}{R+S}\left\{\sum_{j=1}^{R}\sum_{j_m=1}^{j_M}x_{i,j_m}+\sum_{l=1}^{S}x_{il}\right\}$$
$$=\frac{1}{R+S}\left\{\sum_{j=1}^{R}\sum_{j_m=1}^{j_M}x_{jm}\delta_i(j_m)+\sum_{l=1}^{S}a_l u_{il}'\right\}$$

これから，前と同様の意味に於て相関比 max. となる x_{jm}，a_l を求めるのである.

$$(R+S)^2N\sigma_b^2=\begin{pmatrix}\boldsymbol{x}\\\boldsymbol{a}\end{pmatrix}'\begin{pmatrix}H\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\\boldsymbol{a}\end{pmatrix}$$

$$\boldsymbol{x}=(x_{11}\cdots x_{1m}\cdots x_{1M}\cdots x_{j1}\cdots x_{jm}\cdots x_{jM}\cdots x_{R1}\cdots x_{Rm}\cdots x_{RM})'$$
$$\boldsymbol{a}=(a_1, a_2\cdots a_l\cdots a_S)$$

ここで H は次の様に matrix の積に依て表される.

即ち，$H=MM'$ ここで M は $(T+S)$ 行 N 列の行列であり，その j_m 行 i 列及び $(T+l)$ 行

i 列の要素はそれぞれ,

$$M_{j_m i}=\delta_i(j_m)-\frac{f_{\cdot j_m}}{N}$$

$$M_{(T+l),i}=u_{il}'$$

$$(R+S)N\sigma^2=\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{a}\end{pmatrix}'\begin{pmatrix}F\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{a}\end{pmatrix}$$

ここに F は対角行列であり,その j_m 行及び $(T+l)$ 行の要素は

$$F_{j_m,j_m}=f_{\cdot j_m}$$

$$F_{(T+l),(T+l)}=\sum_i u_{il}'^2=N$$

即ち $(T+1)$ から $(T+S)$ 迄の行はすべて N, ここで, η^2 を max. にする為には

$$(H)\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{a}\end{pmatrix}=(R+S)\eta^2(F)\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{a}\end{pmatrix}$$

を解けばよい. これから質的要因の各カテゴリーへ与える数量 x_{j_m} 及び量的要因の変換のパラメーター a_l がきまり i なる個のとる質的特性の数量 x_{i,j_m} 量的特性の数量 $x_{il}(x_{il}=a_l u_{il}')$ から個々のものの数値がきまる.

(v) 要因に質的なものと,量的なものとの両方を含む場合 (その 2)

次に, (i) の方法 (項目がカテゴリーに分かれている場合) と (iii) の方法とを併せて数量化をする. これも前記と全く同様に展開出来る. 即ち i なる個に与える数量は

$$\bar{x}_i=\frac{1}{R+S}\left\{\sum_j\sum_{j_m}x_{i,j_m}+\sum_l x_{il}\right\}$$

$$=\frac{1}{R+S}\left\{\sum_j\sum_{j_m}x_{j_m}\delta_i(j_m)+\sum_l a_l\left(u_{il}^2-\frac{1}{N}\sum_i u_{il}^2\right)+\sum_l b_l\left(u_{il}-\frac{1}{N}\sum_i u_{il}\right)\right\}$$

これから前と全く同様の手法に依り,

$$(H)\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{a}\\ \boldsymbol{b}\end{pmatrix}=(R+S)\eta^2(F)\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{a}\\ \boldsymbol{b}\end{pmatrix}$$

なる解を求めればよい. ここに H は次の様に行列の積に依て表される.

$$H=MM'$$

ここで M は $(T+2S)$ 行, N 列の matrix であり,その j_m 行 i 列, $(T+l)$ 行 i 列, $(T+2l)$ 行 i 列の要素は夫々,

$$M_{j_m i}=\delta_i(j_m)-\frac{f_{\cdot j_m}}{N}$$

$$M_{(T+l),i}=u_{il}^2-\frac{1}{N}\sum_i u_{il}^2$$

$$M_{(T+2l),i}=u_{il}-\frac{1}{N}\sum_i u_{il}$$

F は $\begin{pmatrix}F^{(11)} & 0 & 0\\ 0 & F^{(22)} & F^{(23)}\\ 0 & F^{(32)} & F^{(33)}\end{pmatrix}$ なる部分に分かれ,これは各々対角行列であり,

$F^{(11)}$ は $T\times T$ の行列でその j_m 行の要素は

$$F^{(11)}_{j_m,j_m}=f_{\cdot j_m}$$

$F^{(22)}, F^{(23)}, F^{(32)}, F^{(33)}$ はともに $S\times S$ の行列で各々その l 行の element は

$$F_{ll}^{(22)}=\sum_i u_{il}^4-\frac{1}{N}\left(\sum_i u_{il}^2\right)^2$$

$$F_{ll}^{(23)}=F_{ll}^{(32)\prime}=\sum_i u_{il}^3-\frac{1}{N}(\sum_i u_{il}^2)(\sum_i u_{il})$$

$$F_{ll}^{(33)}=\sum_i u_{il}^2-\frac{1}{N}(\sum_i u_{il})^2$$

である．これより質的要因に与える数量 x_{jm}，量的要因の変換のパラメーター a_l, b_l が求まり，量的要因に新たに与える数量 x_{il} がきまる．かくして，個に与える数量も $x_{i,jm}, x_{il}$ の平均値から求まる．

(vi)　計算法の問題に少し触れよう．上述の (iii), (iv), (v) の方法では，H, F の行列は，夫々，$2S$, $(T+S)$, $(T+2S)$ 等で大変大きな行列となる．それでももしも N がこれ等の数より小であるならば——一般に N はサンプルの個体の数であるから N の方が遙かに大である場合の方が常識的であるが，ここで用いたイネの分類の様に $N=32$, $S=20$, $T\geqq 57$ で，$2S$, $(T+S)$, $(T+2S)$ 等が N より大なる事もある．——次の様な方法を用いた方が便利である[10]．即ち上記 (iii), (iv), (v) に於ける matrix H は何れも $H=MM'$ なる型であるから固有方程式は次の様に書ける．

$$MM'\boldsymbol{x}=\lambda F\boldsymbol{x}$$

ここで $\lambda=c\eta^2$ (c は上記 (iii), (iv), (v) に於ける η^2 にかかる常数) とおく事とし，$\boldsymbol{x}$ は求めるベクトルを表わす事とする．

これは $M'F^{-1}M\boldsymbol{z}=\lambda\boldsymbol{z}$ を解く事と等価である．而して $\boldsymbol{x}$ は，$\boldsymbol{x}=F^{-1}M\boldsymbol{z}$ として求まる．こうすれば N 元の固有方程式を解けばよく，N が小さいものの場合は比較的簡単である．——然も F は特殊な型であるので F^{-1} は簡単に求まるから，$M'F^{-1}M$ を作る事は容易である——．

[実際の適用]

扨，ここで上記の方法を適用してイネの系統の位置づけを行ってみた．尚，ここに用いた対象は No. 26, No. 34 を除いた 32 系統である．——最初の形質間の相関図を眺め，又 I(ii) に依ってもこの2系統は他と異質なる事を知り，且この系統は Oryza 属と考えぬ方がよいかも知れないという森島啓子氏（国立遺伝学研究所）の意見もあったので以下の分析ではこの2系統は除く事とした——．

(1)　定性的要因について

方法の (i) に述べた事を適用した．即ち，質的形質 (21～42) の中，29, 30 を除き（これ等の形質は系統 No. 26, No. 34 の2系統を除いた時，全系統がすべて同じカテゴリーとなって了うから），20 個の形質を用いてこの中の各カテゴリーの反応パタンから 32 系統の位置づけを行うものである．

固有値を第2根迄求め，各々の固有ベクトルから形質のカテゴリーに与える数量を，又その反応した形質のカテゴリーの平均値から系統の値を求めた．${}^1\eta$, ${}^2\eta$ に対応する系統の数値を夫々 ${}^1y$, ${}^2y$ とし，これを2次元に図示すれば次の如くである．以後，図に記入してある ${}^1\eta, {}^2\eta$ は相関比であり，夫々，第1の，第2の分離の尺度といえよう．尚，以後図示に於ける数値は，すべて形質に関する total variance を1になるように変換した数値である．従って縦軸，横軸に夫々投影した時の標準偏差，即ち系統間の標準偏差は，記入してある η と等しくなっている．この図によれば，${}^1y$ により，系統 No. 1～No. 16 とそれ以外とが分かたれ，${}^2y$ に依てこの中，更に No. 1～No. 10 の集団と No. 11～No. 16 の集団，及び No. 31, 32, 33 等が孤立し，No. 17～No. 30 等が一団をなしている事が解る．この時形質に与える数量を項目別にカテゴリーに分けて図示すれば次の様である．

---

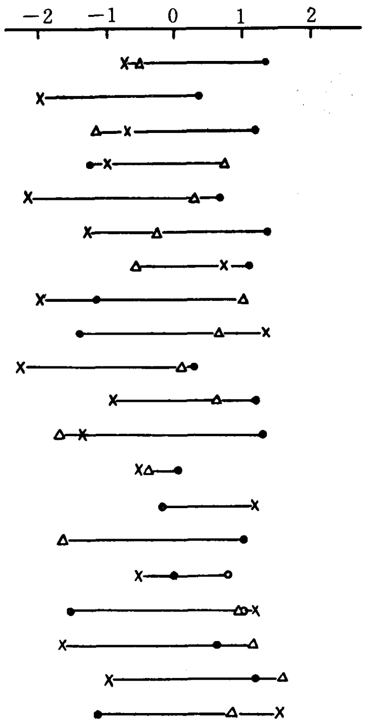
10)　H. Akaike: On a Computation Method for Eigenvalue Problems and its Application to Statistical Analysis. Ann. Inst. Stat. Math. Vol. 10, No. 1.
T. Uematu: Note on the Numerical Computaion in the Discriminant Problem, Ann. Inst. Stat. Math. Vol 10. No. 2.

形質のカテゴリーに与える数値

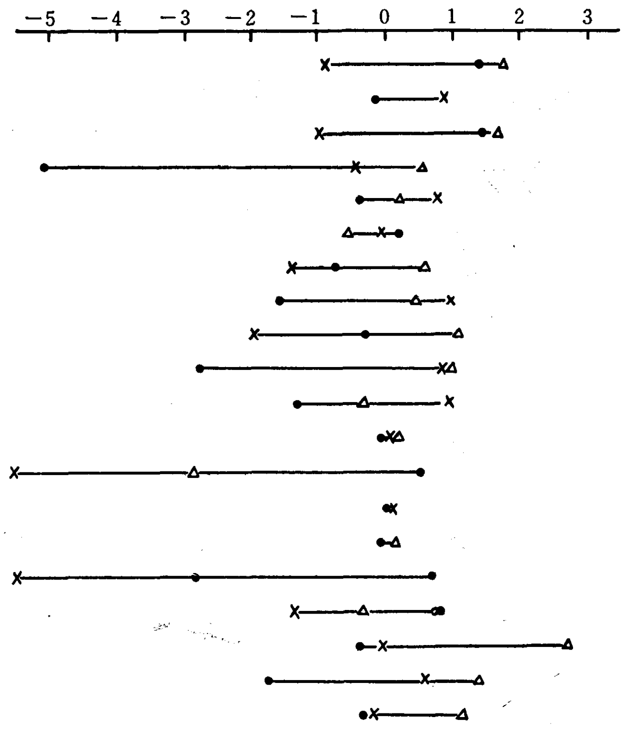
カテゴリー	0	1	2	3
印	○	•	△	×

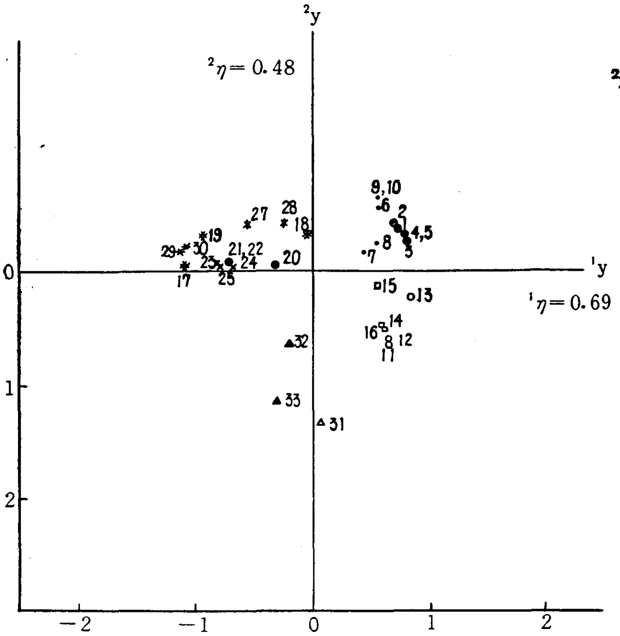
カテゴリーに反応している頻度 0	1	2	3	形質 No.
	11	1	20	21
	27		5	22
	12	1	19	23
	1	19	12	24
	2	26	4	25
	14	4	14	26
	9	20	3	27
	8	19	5	28
	12	14	6	31
	8	22	2	32
	11	4	17	33
	17	1	14	34
	29	1	2	35
	27		5	36
	20	12		37
3	27		2	38
7	12	11	2	39
	20	3	9	40
	9	5	18	41
	16	7	9	42

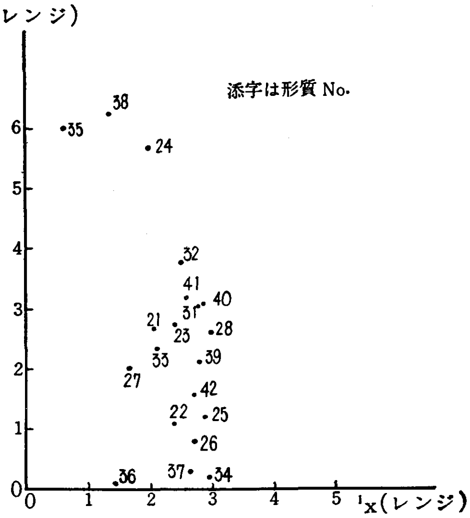
第1根に依る



第2根に依る





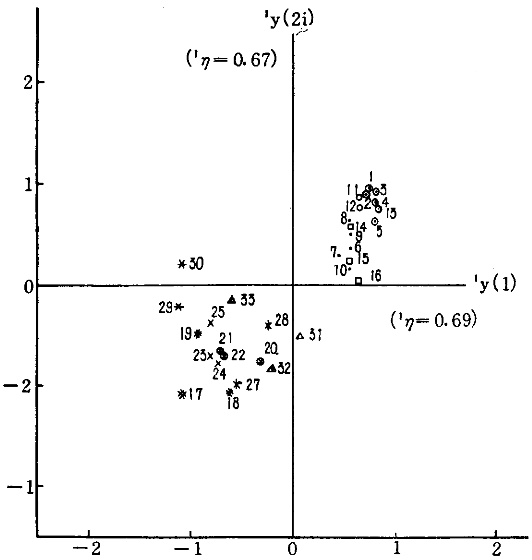


これを見ると各項目内でカテゴリーに与える数値のレンジは全般的に第1根に依る値の方が小さく，且項目別に一様である．第2根による値では項目によりレンジの大きいものもあるが（形質 No. 24, 35, 38）これ等の飛び離れた値は頻度の少い所である．即ち或項目に於て殆ど大半のものが1なるカテゴリーに反応を示しているのに，別の3なるカテゴリーに反応するという事は異質的なものと考えられ，依て非常に離れた値をとるという事もうなづけよう．この様にカテゴリーに反応を示す総数は夫々異なるので，項目内のレンジをとって，これを以てこの項目の分離への寄与であるという事は，一概にいえない事であるが，項目がどの様なきき方をしたかの一応の目安となろう．依て，第1根，第2根に対する項目別のレンジをとってこれを2次元に図示すると上の如くである．これに依ると，第1根に対するレンジではどの形質も略同じ位ではあるが，35, 38, 36, 27 等の形質が他に比べ幾分きき方が少い様に思える．この事は系統の分れ方及び最初のデーターに立ち帰って形質の分布図と勘案した時，納得のゆく事である．即ち，第1根に依って系統 No. 1～16 とそれ以外に分れるのであるから，この弁別力の弱い形質は第1根に対して当然きかなくなっているであろう．然も，これ以外の殆どの形質が大体同じ様にきいている事は，どの形質に於ても系統 No. 1～16 とそれ以外とは，総じて異なったカテゴリーに属する事が多い様である．又 ^2y に依ては系統 31, 33 の孤立，及び 11～15 とそれ以外とのかたまりをなすのであるが，これを弁別するに強い形質 38, 35（系統 No. 31, 32, 33 を弁別する要因である）がきき，形質 36, 34, 37 の様に，系統 No. 1～16 とそれ以外とを分ける力しかないものはきいていないのである．

(2) 定量的特性について

量で測定出来る形質 1～20 を用いて分類を行った．

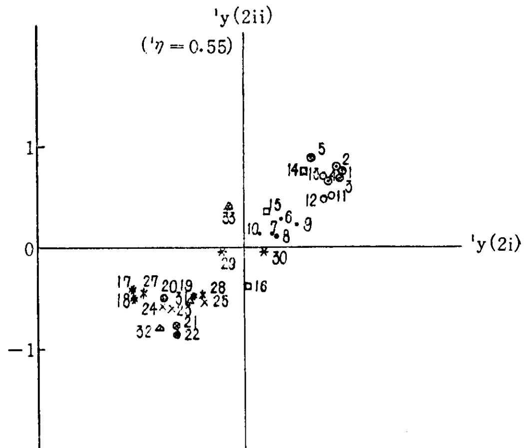
(i) 先づ初めに夫々の特性の量を5つのカテゴリーに区切り（この区切り方は先に述べた e_{ij} を用いた時と同様にした），その各々への反応パタンを作り，これから (1) の定性的な場合と全く同様にして，固有値及び固有ベクトルを計算した（これは第1根だけを算出した）．この相関比は 0.67 で一見よい分離であったかに見えるが，これは形質のカテゴリーを細かくとったからであって——カテゴリーを細かくとればとる程，個別の系統は異った数値をとる事が出来るのだから η の値は当然高くなる.——32 の系統を分けるのに5つのカテゴリーにしたのでは，その各々に属する



ものは平均的に6個程であり（少いものでは1つのカテゴリーに1系統しか入っていないものもある），各カテゴリーに与えた数値は安定した値とはいえないものであろう（この数値は，後に2次式に依る変換の場合の時に一緒に図示する）．然し，これにより，量的形質からする系統の大体の分類は窺えるであろう．この分類に依る系統の数値を ^1y (2 i) とし，前記 (1) の定性的形質に依る 1 根に対する系統の数値 ^1y (1) とを2次元に図示すれば左の如くである．これに依ると，^1y (1)，^1y (2 i) の相関は可成り高く見られ，つまり ^1y (1) に依る分類も，^1y (2 i) による分類も大体同じ事を表はすに過ぎない事が察せられる．

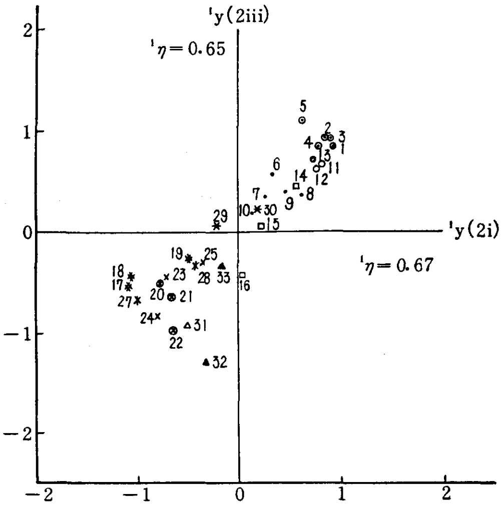
(ii)　次に方法の (ii) を適用してみる．

測定された数量は分類に対して線型の関係を持つと想定したのである．これは各々の形質に対する重みを与えるものであり，1形質に対し1つのパラメーターのみを与えるのであるから，前記の様な数値の不安定さはない．然し，線型的関係で縛って了うので，分離度は悪くなる ($^1\eta=0.55$)，これに依て系統に与えらる第1根に対する数値を ^1y (2 ii) とし，これと前述の量を5カテゴリーに分けた時の系統の数値 ^1y (2 i) とを2次元に図示すれば次の様になる．これに依ると (2 i)，(2 ii) の方法に依る分類も分離度は多少異なるとしても，略々同じ結果となる事が知れよう．尚，この場合の形質に与える重み，即ち1次変換の様子は，後述の2次変換の場合と同時に図示する．



(iii)　次に方法の (iii) を適用する．

(2 i) に依る 5 つのカテゴリーに区切った場合の形質の数値から推察するに（後に示す形質に与える数量の図参照），量的特性は系統を分類する上から見てすべての形質が線型的関係にあると仮定する事は少し無理の様である．依てその測定値を形質毎に2次式で変換し，然る後に，その数値を分類に用いてみては如何であろうか．それで前述の方法 (iii) を適用してみた．これにより第1根を計算すると $^1\eta=0.67$ となり，形質を5カテゴリーに分けて得た数値に依る分類と殆ど同じ効果を挙げている．即ちこの2次式に依る変換は，この場合には非常によく適合したものといえよう．これに依れば，各形質に与えるパラメーターは2つで新たな形質の値が決定されるのであるから，その数値は，5カテゴリーに区切った時に比べ，遙かに安定した数値となろう．これに依る系統に与える数値を ^1y (2 iii) とし，これと ^1y (2 i) とを夫々軸にとって図示すれば次の如くである．この図から見ても，この2方法により系統に与えた値は，非常に相似たものであり，より少ないパラメーターでつまり安定した数値で同じ分類を行う事が出来た事は，この方法が極めて適切であったといえよう．尚，この場合の形質に与える数値も，5カテゴリーに分けてその各々に与える数値と極めてよく似たものであり，次図に依て見られる如

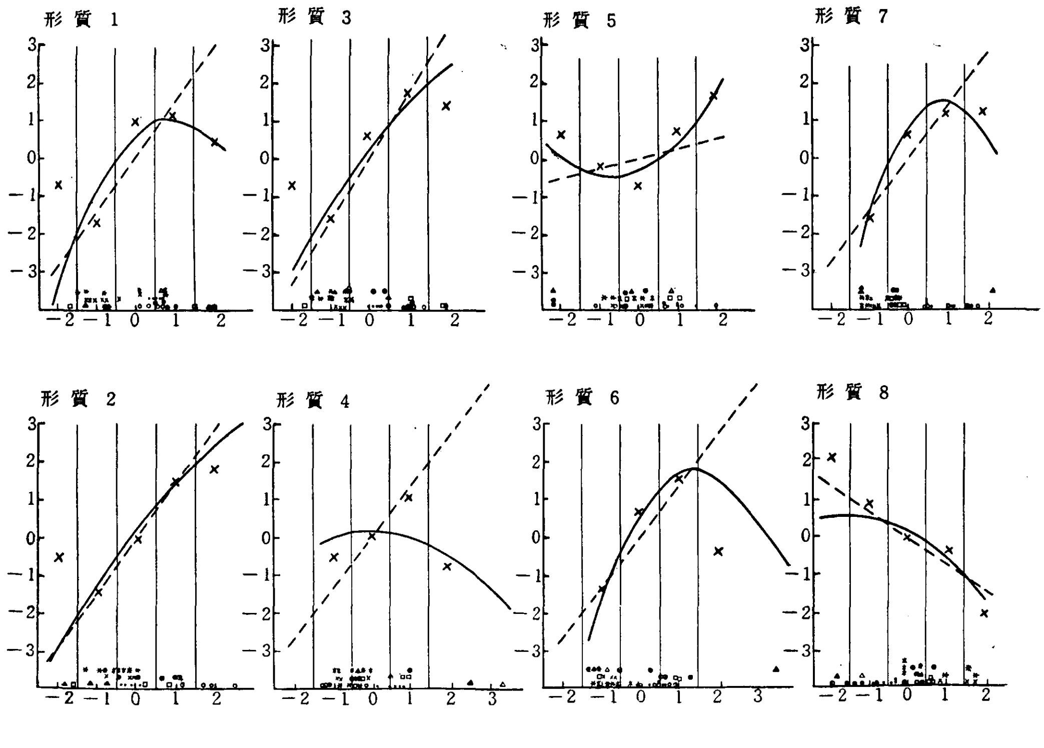


く恰も5カテゴリーに与えた数値を最小二乗法で結んだかの様である．

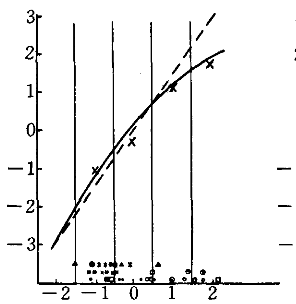
以上，量的形質を用いて行った3つの分類法に依て，形質に与えられた数値を綜合して図示すれば次の如くである．何れも第1根に対する数値である．

（註） 系統測定値の印と系統 No. との対応

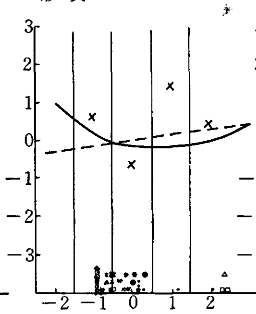
印	系統 No.	印	系統 No.	印	系統 No.
○	1～5	×	17～19	⊗	29～20
・	6～10	＊	20～22	△	31
◎	11～13	＊	23～25	△	32
□	14～16	✳	27～28	△	33



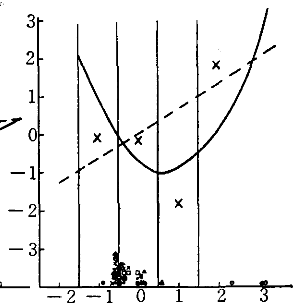
形 質 9



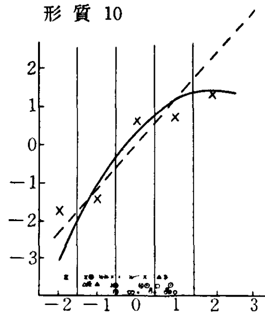
形 質 11



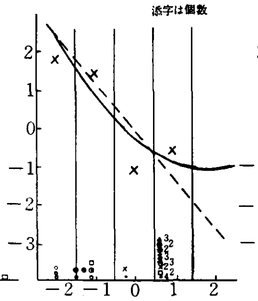
形 質 13



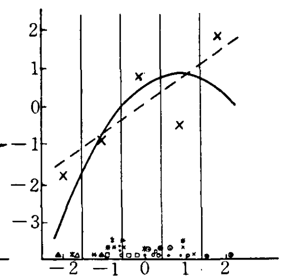
形 質 10



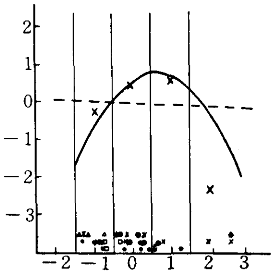
形 質 12



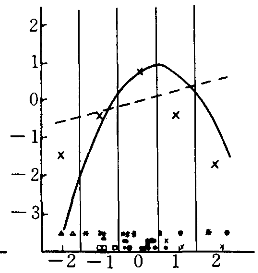
形 質 14



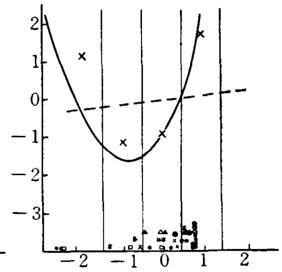
形 質 15



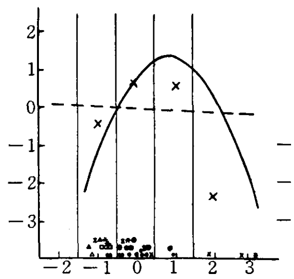
形 質 17



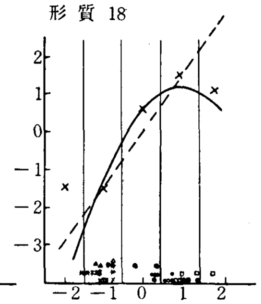
形 質 19



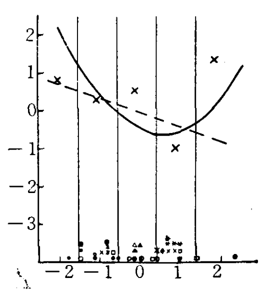
形 質 16



形 質 18



形 質 20



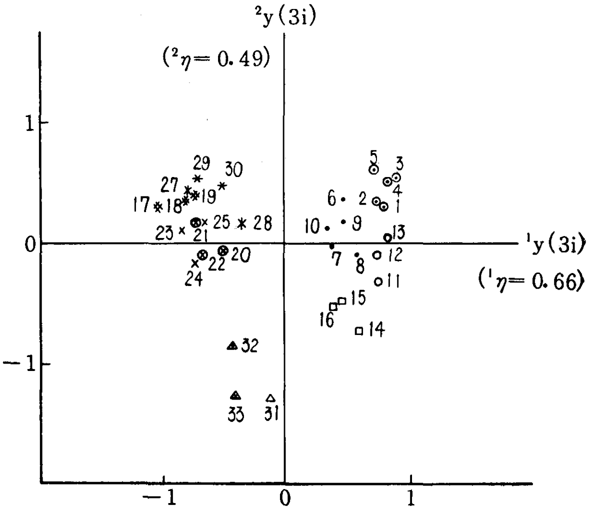
即ち，X 軸について目盛ってある諸印（分布）は，系統に関する始めの測定値（平均 0，分散 1 に変換してある）である．これを5つのカテゴリーに区切ってその各々に属するものに新たに与える数値は×印で記されてある．ここに引かれてある直線は（2 ii）の方法により1次変換を行った時の関係である．即ちこれに依て最初の測定値は対応する Y 軸上の値となり，これが与える数値 ${}^1x(2\,\mathrm{ii})$ となる．曲線は（2 iii）の方法に依る2次式に依るものである．即ち，最初の数値をこの2次曲線にのせることに依て，2次変換後の与えるべき数値となる（この2次曲線と×印とが非常によく合っている事が解るであろう）.

(3) 定性的特性と定量的特性との併用

量的特性に依る分類と質的特性に依るそれとはよく似た様相を呈しているので——例えば ${}^1y\,(1)$ と ${}^1y(2\,\mathrm{i})$ とは相関が高い——これを併せて同時に行っても，余り分離度はよくならないと思えるが，兎も角測定出来た 40 形質を綜合的に用いてみたかったので，一応併せて行って見る事とした.

(i) (1) と (2 i) との併用

これは量的特性をも質の場合と同様に扱ったのであるから，特性が増したというだけの事である．これに依り第1根，第2根とを計算し，これに対応する系統の位置づけを 2 次元に図示した．これに依ると，No. 1～No. 16 とそれ以外とに大別され，No. 31，32，33 が孤立するという顕著な様相は質的特性のみで行った場合と同じである．但し No. 1～No. 16 の中は多少異なり，No. 1～No. 5，No. 6，9，10 は近く，No. 14～No. 16 は稍々離れている様に見られるが，はっきりしたものではない.



(ii) 方法（iv）の適用

質的要因では反応したカテゴリーの数量を，量的要因は1次変換を行った数量をとり，その平均をとって系統の値としたものである．第1根，第2根に対する数値による位置づけを図示してみたが，概要は前記と略々同じ様相を呈した.

(iii) 方法（v）の適用

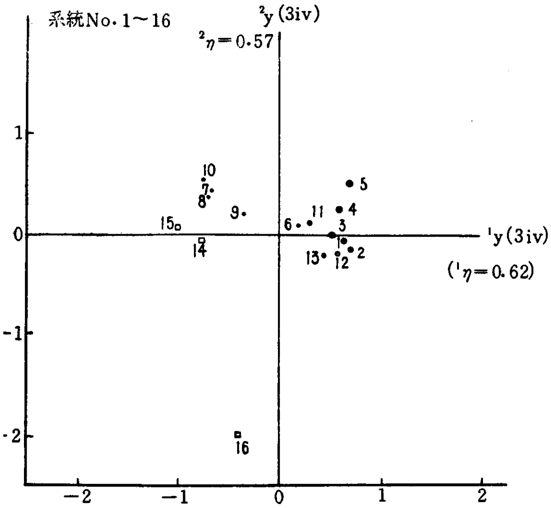
質的要因では反応したカテゴリーの数量を，量的要因では2次変換を行った数量をとったものである．これに依り2根に対する系統の位置を図示したが，これもはっきりした様相を示す限りでは（i）の場合と殆ど同じであった.

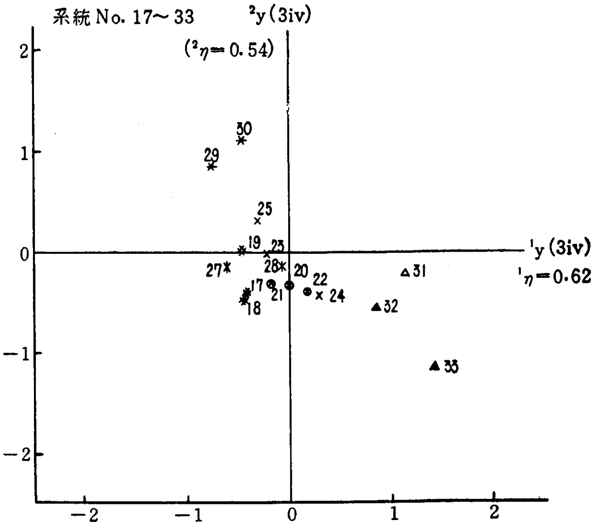
要するに，質的特性と量的特性とを併せて 40 形質に依る分類を行ってみても，質的特性のみによるものと，顕著な様相の限りに於ては大体同じ様なものである．即ち，以上の何れの方法に依っても，はっきりしている事は No. 1～No. 16 とそれ以外とは明らかに別のグループをなすという事である．依て，次にはこれ等の2つのグループを別個に扱って見ることとした.

(iv) 質的特性で分かれたサブグループに於ける量的要因に依る分類

質的特性を用いて分類した場合に（量的特性を加えても同様），系統 No. 1～No. 16 とそれ以外とが大別される事は明らかであったので，この2集団を別々に取扱い，その各々に量的特性に依る分類を行ってみた．この場合，量的特性は3つのカテゴリーに分けてこれに数値を与える方法をとった．カテゴリーを3つにしたのは，分類すべき系統が各々 16 づつであるから，前の場合の様にカテゴリーを多くとると，1つのカテゴリーに反応するものは少くなり，与える数値は不安定な

のとなるからである．この区切り方は，形質の分布の型等から考えて，(−0.5未満)，(−0.5〜0.5)，(0.5以上）とした．これにより，夫々第2根迄を計算して2次元に図示すれば次の如くである．





これに依ると，No. 1〜No. 16 に於ては，先に 2 (i) 即ち量的特性を5カテゴリーに分けて全系統について行った分類の，第1根に依るものと特に似た位置づけであり，即ち，No. 1〜5，No. 11〜13，No. 6 が近くに集まり，No. 7〜15 は稍々離れている．唯，No. 16 が第2根で離れたという事だけが，顕著な事である．No. 17〜No. 33 の方では，やはり No. 31，32，33 は孤立し，No. 29，30 は近くにあり，他は略々一群をなしていると見られ，前述の 3 (i) 即ち“質的形質と，5カテゴリーに分けた場合の量的形質とを併せて行った方法”に於ける第2根に依る分類と相似た様子である．

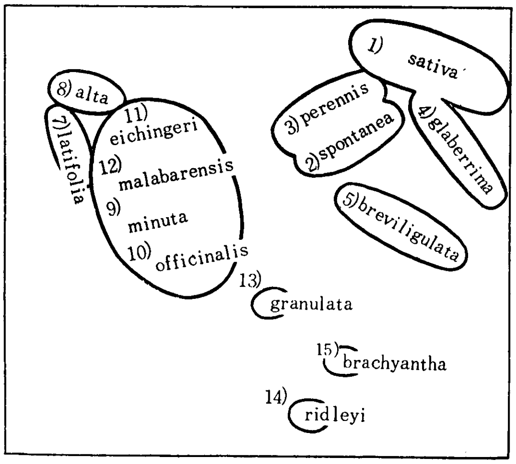
(4)　数量化法を適用した場合の分類を綜合して

以上，これ等の形質に依る位置づけの方法を綜合して考察するに，始めに述べた如く，生物の分類には，所謂外的規準となるものがないので，分類方法の良否については常に絶対的判断をする事は出来ない．この分析の結果から見ると，旧来同一種とされていたものの数系統は互に近くに位置したこと，又種間の距離が従来分類学者が考えていた所とよく似通っている事（Roschevicz (1931) 森島・岡 (1960) 等参照）等が指摘される．然し同一種からとられた系統の数が不揃であった為，多数系統を有する種 (sativa, spontanea) を弁別する特性に多くの重みがかかって了ったという憂もある様に思える．

兎も角，数量化法により，40 の形質からする系統の位置づけを考えると次の様である．

(i)　個々の系統は，やはり旧来それが所属している種に依てグループされ得る．

(ii)　種の相対的位置を極く大ざっぱにいうならば，次の様に図示されよう．これは本報告における種々の方法に依る分類結果を模式的に綜合して描いたものである．尚，ここには系統 No. 26 ([6] australiensis) と No. 34 ([16] subulata) とは



（註）種名に付したsuffixは種No.である

除いてある．これは前述の様に，系統に対して安定した要因である定性的特性に於て，他の系統と著しく異って居るので，別個に考慮する事とする．尚，perennis と spontanea との差異についてはいろいろ意見があるが，このデーターに依る分類程度の粗さでは，両者は殆ど区別出来ない種の様に見られる．

§5. 森島・岡によるイネ属の分類について

ここに用いたのと同一のデーターを用いて，種を単位とした分類を遺伝学研究所の森島・岡(1960) が行っている．これには次の２つの方法が用いられた[11]．

i) Sokal and Michner の Weighted variable group method に依る方法

種間の相関行列に基づき，相関係数の高い２つの種を核とし，これに順次高い相関のものを，或方法に従って順次つなぎ合せて樹枝状に図示するものである．これに依ると，極く大まかに見れば種 No. {1～5}，{7～13}，{14～15} 等のグループに大別されるようである．

(ii) 因子分析による方法

averoid solution に依り（これは重心法に於てコミュナリティを定めるのに，相関行列の各行の平均値をとったものである)，第 4 因子迄を抽出し，これ等４つの因子の中２つの因子を直交軸にとり，夫々の因子負荷に依り 16 種の位置づけを行ったものである．これに依ると，第１因子と第２因子に依る２次元の図からは，やはり Sokal の方法の場合と同様，種 {1～5}，{7～13}，{14～15} 等のグループに大別されている様子が見られる（但し "autraliensis" と "subulata" の位置はこの２法に於ても一定していない)．

これ等の結果を見ると，数量化で行った結果と比べ，種 No. 13 (granulata) の位置が少し違う他は，粗いグループ分けに関しては大体相似たものとなっている．然し，この方法に就て意見を述べるならば，次の事がいえよう．

第１にこの２つの方法は，どちらも系統間の関係を相関係数の尺度で測り，これから出発している．ここで先づ問題になるのは定性的形質についてである．これは数量で測定出来ない形質であり，便宜上のコードをカテゴリーに付しただけのものである．依てこのコードを数量として扱って相関係数を計算することは不合理であろう．且つ，この解析のＩに述べた様にコードは任意につけられるものであり，然もこの符号のつけ変えに依て相関係数は値が変って了うのであるから，この関連の尺度は少し不安定なものといえよう．次に（定量的特性をとり上げたとしても）相関係数は，系統間の関連をリニアーな関係に於て測る尺度であり，この場合には，関連について１次の関係に縛らぬ方がよりよい効果がある様である（前述 II 2 (i), (iii) を (ii) と比べる事に依て察せられる)．

次に，この２方法の各々に就て次の事がいえよう．先づ，Sokal の方法は，基本としては因子分析法に於いて予め変量（ここでは個体）をグループする時に用いる B-coefficient の尺度に基づくものと見られるが，この方法は計算結果から直ちに位置づけが決まるわけではなく，この尺度を目安として手内職的に樹枝状図につないでゆくものである．依て稍々主観的な判断を含むものであるのでこれを客観的な方法に発展させたのが，§4 I に述べた e_{ij} 型の数量化である．

次に因子分析法に就ての考察を少し述べよう．因子分析法に於て行はうとする事は，次の様な事であろう．即ち，先づ或変量（心理学の立場からいうならばテストという）が潜在的な因子（共通因子と特性因子）の一次結合で表現されている事を想定する．そして，ここで行われた重心法にせよ，主因子解にせよ，その様な解法はすべてこのテストに対し，共通であると考えられる因子の負荷を，その共通性の強い順に算出する事である．——普通第１因子，第２因子，等といわれているものはこれ等の因子負荷の事である．——依て算出された因子負荷はあく迄最初のモデルに於ける

11) H. Morishima, H. Oka: The Pattern of Interspecific Variation in the genus, Oryza: Evolution Vol. XIV, No. 2

因子のウェイトであって，これはテストを分類する目的でなされた座標となるものではないのである．唯，仮想せる同じ因子に対して同じ様な負荷の値をとったものは，似たテストと考えてもよいであろうという想定を入れた上で便宜的に分類に用いているに過ぎない．然も，この方法に依れば，最初のモデルに特殊因子を想定するので（これを省き共通因子だけの一次結合のモデルを考えるのでは因子分析法の味はいが薄れよう），コミュナリティの問題が生じる．これは，次の様な事である．即ち，因子負荷を計算するには，コミュナリティに何らかの推定量を与えねばならない（これを推定する方法はいろいろあるが，筋の通った推定はない様に思える）．然も，この推定値の下で因子負荷の計算を行い，この因子負荷に基づき再びコミュナリティを計算してみると，始に与えた推定値とは概ね異った値となる．この新しい値を第2次近似値として与え，繰り返し計算を行う事は非常に煩雑であり，然もその結果一定の値に収斂するという事は保証出事ない事である（もしこの様な繰り返し計算を行わず，唯一回のコミュナリティの推定量を与え，これから因子解を解いただけでは，因子分析法の理論からして不完全なものといえよう．この様に理論と実際の計算とには若干のギャップがある様に思える．然もこの方法の目的は，テストに対し共通であると考えられる潜在的な因子の負荷を計算してみる事であるので，もし最初からテストの分類を行いたいならば，その事を目的とする尺度をとり上げ（相関比等），これを max. にする方法を用いた方が筋の通ったものとなろう（これが前述の数量化の方法である）．因子分析に於てコミュナリティを 1 にした場合（共通因子のみを考える）と，数量化に於て要因を1次変換して分類を max. にする方法（II (ii)）とは，式の型に於て一致したものであるが，数量化で行はうとしている事は，あく迄直接に分類を考え，これを出来るだけ判然としようという立場から行っているので，この観点に立っていればこそ，要因（これは潜在的な因子ではなく測定値を得ているものである）のきき方に種々の考慮を加えること（例えば量的なものであれば2次変換を行う等）が出来るのであって，その方がより目的に適った方法といえるのではないかと思う．

§6.　終　り　に

以上の方法を省みて，今後の課題となる事は次の事であろう．

素材としたイネ属の分類に関しては，より詳しい結果を得る為には先づデータに考慮を加えなければならない．即ち，系統を選び出す際には，旧来同一種とされているものの中から同じ割合で成る可く多数系統を選び出し，又各系統の個体は，出来るだけ条件を揃え，且つ多くの個体について測定する必要があろう．かくして統計的処理に充分堪え得る確実性をもったデーターから出発すべきであろう．更に要因に就て考えるならば，ここに用いた 42 形質の測定値は，この植物群に含まれる莫大な遺伝的形質の極く一部分を示すに過ぎない．従って，これだけの知見から生物の変異を正しく位置づけることが出来るかどうかという点に危虞もある（遺伝学研究所岡彦一氏に依る）．然し又別の見方をするならば，この報告に於て示された様にとり上げた 42 形質全部を用いた様相も，定性的特性 20 形質のみを用いた様相も殆ど変らないという事から，ここで用いた形質は可成り安定したものであると考える事も出来よう．このためには，困難なことでもあるかもしれないが測定さるべきディメンションとしての形質のユニヴァースを明確に規定し，それから母集団を構成し，層別抽出等の配慮の下に，なるだけ少数の形質の測定によって母集団としての全形質を代表する様に考へるべきであろう．そして更には同じ分類を呈するのに，より少い形質を以て為し得ると考え，この要因となる最小の形質を見出す方法を編み出す事も統計的方面の課題となろう．――この事は生物の分類の場合に限らず一般に社会現象の類型化の時にも必要となる事であろう．――要するに，先に述べた外的規準を決める事にせよ，要因を増減する問題にせよ，斯学との十分の交流を行った上で解決を図りたいものと思う．

この報告をなすに至ったのは，師林知己夫氏の懇切なる御指導のお蔭である．ここに深く謝意を

表する次第である．尚，遺伝学研究所の岡彦一先生[12]からは，この報告について種々御親切な助言を頂いた．深く感謝の意を表するものである．

統計数理研究所

12) 氏の御意見に依ると，「生物の分類に於ては，ここに云う“外的規準”を斯学の中で規定する事は不可能であり，寧ろこの様な静的分類が生物学の進歩への一助となろうと考えられ」「従来分類学者が数十年を要して到達した程度の結論を統計的処理に依っても容易に得られるものである」との事であった．