日作紀（Jpn. J. Crop Sci.）78（3）：395―398（2009）

# 教育のための手計算によるクラスター分析

吉田智彦[1)]・Anas[1)]・小林俊一[2)]
（[1)] 宇都宮大学農学部，[2)] 栃木県農業試験場）

**要旨**：生物種あるいは品種間の相互関係を表示するために，通常はコンピュータソフトを利用したクラスター分析により樹状図を作成しているが，教育的効果を目的としてコンピュータを用いず手動でクラスター分析をすることを試みた．オオムギ品種間の RAPD 分析による DNA 多型データを用いて，品種間で異なるバンドを示した DNA マーカー数（異なるマーカー数）をその品種間での距離とした．まず，異なるマーカー数の最も少ない組合せを選び，それを最初のクラスターとした．次にそのクラスターの平均値からの距離と残りの品種との間の値を計算し直して，第 2 のクラスターを決定し，順次同様に行っていった．育成地の異なるオオムギの二条，六条種を含む品種間で試みたところ，ほぼ満足すべき結果が得られた．コンピュータソフトを利用した結果とも一致した．本方法では，クラスター分析を手計算で行うことにより，理解が容易であり，教育的効果が大きい．
**キーワード**：教育，クラスター分析，樹状図，手計算．

多くの生物種あるいは品種をなんらかの手段で分類し，その結果を系統的に表示してそれら相互間の関係を探りたいことがある．そのときの手法としてクラスター分析がよく行われる．クラスター分析とは，ある集団内の個体をいくつかの似た者同士の群に分類するとき，似た者同士を集める手法の一つであり，似ている程度を測る物差しとしては，各個体について計測された複数の特性値から計算した多次元空間内の距離を通常用いている．従って，分類が恣意的に行われるのでなく，数理的な基準によってなされるので，分類した結果を万人に納得してもらい易いと言える．似た手法として別に判別関数法があるが，これは群が予め設定されているのに反して，クラスター分析では事前の情報なしに距離だけをもとにして似た者同士を集める計算を行い，正規性や線形性の仮定は不要であり，異常データの検出も可能である（奥野ら 1971）．

実際のクラスター分析では，それら対象とする生物種あるいは品種について多くの形質を計測し，それらの値をもとに多次元空間内の距離を生物種あるいは品種相互間で計算し，その距離の最も近いものを第 1 のクラスターとして併合し，併合したクラスターを含め残りについての距離を再計算して次のクラスターを決定し，順次それ以降も同様な計算を行ってクラスター分けを行って，最終的な分類を樹状樹の形にして作図を行う．

また，併合するときの“近い距離”をどう定義するかの違いで，クラスター分析にはいくつかの手法がある．クラスター間の距離をクラスター内の平均値間，最も近隣，あるいは最も遠隣とするかで，メディアン法，最短距離法，あるいは最長距離法，さらに群内個体数による重み付けを行うウォード法などがあり，この中でウォード法は分類感度が高く，最も明確なクラスターを作るとされる（青木 2009）．

いずれの方法にせよ，これらの計算や作図は手計算では不可能であり，コンピュータソフトを用いて行うのが一般的であるが，計算手順や理論に全く触れることなく計算ソフトのみを使って結果を出すことは初学者にとって教育的でないし，ある場合は結果の解釈が不適当になる危惧もある．そこで，ここではクラスター分析を手計算で行うことで理解を深め，教育的効果を高めることを目的として以下を試みた．

## 材料と方法

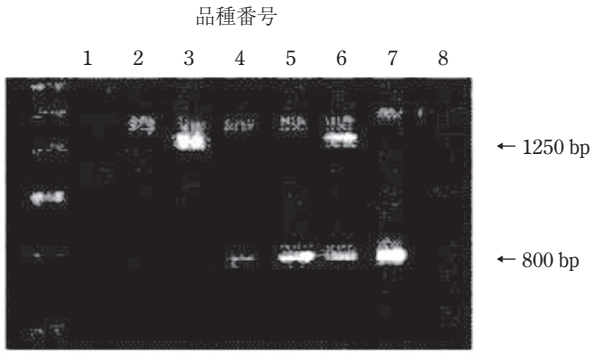
使用したデータはオオムギの品種分類を目的としてなされた，小林・吉田（2006）の一部を抜粋して用いた．ビール醸造用の二条種を 3 品種，食用の六条種を 5 品種供試した．育成地は酒造会社，国や県の試験場など様々になるようにした．

第 1 表に品種名，条性，育成地を示す．第 1 図に，プライマー OPD12 によるオオムギ 8 品種の RAPD 分析による DNA 多型を示す（原図から該当品種だけの部分に加工し

第 1 表　供試品種とその条性，育成地．

| 番号 | 品種名 | 条性 | 育成地 |
|---|---|---|---|
| 1 | サチホゴールデン | 二条種 | 栃木農試 |
| 2 | あまぎ二条 | 二条種 | キリンビール |
| 3 | なす二条 | 二条種 | キリンビール |
| 4 | シュンライ | 六条種 | 長野農試 |
| 5 | カシマムギ | 六条種 | 農事試 |
| 6 | マサカドムギ | 六条種 | 農業研究センター |
| 7 | ファイバースノウ | 六条種 | 長野農試 |
| 8 | イチバンボシ | 六条種 | 四国農試 |

2008 年 11 月 25 日受理．連絡責任者：吉田智彦　〒 321-8505 栃木県宇都宮市峰町 350
TEL 028-649-5429，FAX 028-649-5401，yoshidat@cc.utsunomiya-u.ac.jp




第1図 プライマー OPD12 によるオオムギ8品種の RAPD 分析による DNA 多型.

第3表 品種相互間で異なるマーカーの数.

| 品種番号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 10 | 6 | 22 | 23 | 21 | 19 | 15 |
| 2 | | 0 | 10 | 20 | 21 | 27 | 21 | 19 |
| 3 | | | 0 | 20 | 21 | 19 | 17 | 13 |
| 4 | | | | 0 | 11 | 15 | 5 | 15 |
| 5 | | | | | 0 | 8 | 10 | 10 |
| 6 | | | | | | 0 | 12 | 10 |
| 7 | | | | | | | 0 | 10 |
| 8 | | | | | | | | 0 |

第2表 オオムギ品種の RAPD 分析による DNA 多型.

| 品種番号 | 33マーカーのバンド有無（1，0で示した） | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | **0** | **0** | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 2 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | **0** | **0** | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 3 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | **1** | **0** | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 4 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | **0** | **1** | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| 5 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | **0** | **1** | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| 6 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | **1** | **1** | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| 7 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | **0** | **1** | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 8 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | **0** | **0** | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |

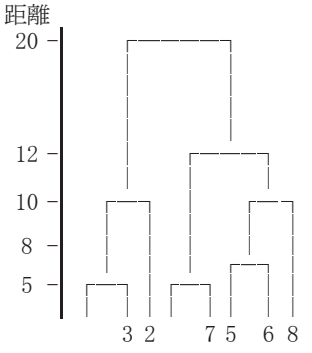
データは小林・吉田（2006）から一部抜粋した．品種番号は第1表を参照.
太字は第1図に示したプライマー OPD12 による 1250 bp と 800 bp のバンドの有無である.
他のプライマーについては，小林・吉田（2006）を参照.

た）．バンドの検出されたものを1，検出されなかったものを0とした．他のプライマーについても同様に行った（他の図は省略）．全33マーカーの結果を第2表に示した．太字は第1図に示したものの結果である．第2表の値について，品種相互間で異なるバンドを示した DNA マーカー数（異なるマーカー数）を数え，第3表に8品種相互間の値を示した．例えば，品種1と品種2の間では，異なるマーカー数は10である.

品種間で異なるマーカー数は，その品種間がどの程度離れているか，つまり品種間の距離を示すと考えられるので，この値を用いて以後の計算を行うこととした.

なお，数の多い品種相互間での異なるマーカー数の計算は煩雑である．この計算は筆頭筆者の Web サイト（吉田 2009）内のユーティリティ使用で簡単に計算可能であるが，少なくとも2，3品種間は実際に数えさせるほうが教育的であろう.

また，ここでの手計算結果との比較のためにクラスター分析をコンピュータソフトで正式に行って比較した．計算ソフトは前述の青木の Web サイト（青木 2009）のものを用いた.




第2図 二・六条品種を含めた結果.

## 結果と考察

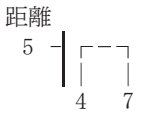
練習用に，六条種だけで計算してみる．第4表に六条種についての値を示した．ここで，一番“近い”のは4対7の5（太字）である．従って品種番号4，7を1つのクラスターとし，その距離を5とした（第4表の第1段階）.

次のクラスターを決めるため，第1のクラスターからの平均値を計算する（つまり“メディアン法”をとることに

第4表　手動によるクラスターの計算.

第1段階

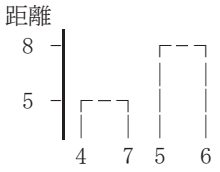
| 品種番号 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|
| 4 | 0 | 11 | 15 | **5** | 15 |
| 5 | | 0 | 8 | 10 | 10 |
| 6 | | | 0 | 12 | 10 |
| 7 | | | | 0 | 10 |
| 8 | | | | | 0 |




第2段階

| 品種番号 | (4, 7) | 5 | 6 | 8 |
|---|---|---|---|---|
| (4, 7) | 0 | 10.5 | 13.5 | 12.5 |
| 5 | | 0 | **8** | 10 |
| 6 | | | 0 | 10 |
| 8 | | | | 0 |

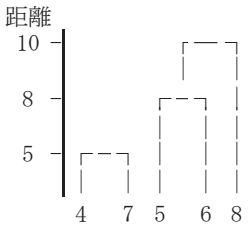
例；10.5=（11+10）/2




第3段階

| 品種番号 | (4, 7) | (5, 6) | 8 |
|---|---|---|---|
| (4, 7) | 0 | 12 | 12.5 |
| (5,6) | | 0 | **10** |
| 8 | | | 0 |

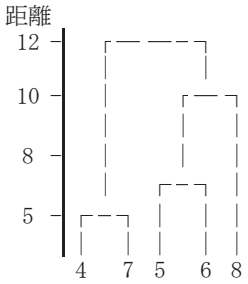
例；12=（10.5+13.5）/2

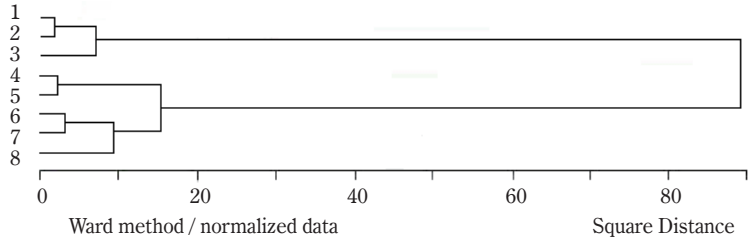



第4段階

| 品種番号 | (4, 7) | (5, 6, 8) |
|---|---|---|
| (4, 7) | 0 | **12.25** |
| (5, 6, 8) | | 0 |

例；12.25=（12+12.5）/2




第3図　コンピュータソフトによるクラスター分析の結果.

なる）．例えば（4，7）と5の距離は（11+10）/2=10.5であり，他の組合せも同様に計算すると5対6の距離が最小で8である．よって，第4表の第2段階のように次のクラスターが決定される．

次に，（5，6）との平均距離を計算すると，（5，6）対8が最小で10である．よって第4表の第3段階のようにクラスターが決定される．

次は，（5，6，8）と（4，7）の距離が12.25となり（第4表の第4段階），六条種だけの類縁関係ではあるが樹状図が完成する．

この分析により，六条オオムギ品種が育成地別（農事試と農業研究センターの関東と長野，四国）に手計算で分類できた．なお，4，7（長野）と5，6，8（関東，四国）が異なる群に分類されることも示された．

例示は六条種品種だけにして，以後は学生への宿題として二条種を含めた全品種について計算することを課題とすると，より理解が深まる．

二条種を含めた結果を第2図に示した．二条種と六条種のオオムギは用途，導入の過程が元々異なっており，遺伝的背景が大きく異なることは既に知られている（増田

1993）．手計算による簡単な本解析でもそれを裏付ける結果が得られ，これにも教育的効果が期待できる．

また，クラスター分析をコンピュータソフトで正式に行った結果を第3図に示す．ここではデータを標準化し，ウォード法をとっている．手計算で行ったものと同じ結果である．

従来は，DNA多型の解読の段階からイメージスキャナで自動的に行い，その数値解析も付属の計算ソフトで連続的に行うので，完全にはそれらの操作を理解せずに行っている場合も多いと推察される．ここで示した方法では，DNA多型を判読し，その値を使って初歩的な計算を行うことでクラスター分析の概要を容易に把握することができる．類似の問題について以後たとえコンピュータソフトを使うにせよ，一度手計算を経験しておくことは，クラスター分析への理解を深めるのに極めて有益と考えられる．実際に大学院の講義の一環として本方法を導入したところ，すべての学生がクラスター分析への理解が深まったとの感想を述べた．また，講義として受け身で聞くのみでなく，二条種を含めた全品種での計算を宿題として独力で解く過程が理解を深めるのに必須であった．最終試験で類似の問題を提出したところ，大部分の学生が正しく計算と作図をした．

## 引 用 文 献

青木繫伸 2009. http://aoki2.si.gunma-u.ac.jp/index.html（2009/4/14 閲覧）.

小林俊一・吉田智彦 2006. RAPD分析による栃木県を中心とした関東周辺地域のムギ類優良品種識別. 日作紀 75：165－174.

増田澄夫編 1993. わが国におけるビール麦育種史. ビール麦育種史を作る会, 東京. 1－452 .

奥野忠一・久米均・芳賀敏郎・吉澤正 1971. 多変量解析法. 日科技連, 東京. 1－430.

吉田智彦 2009. http://www.d1.dion.ne.jp/~tmhk/yosida.htm（2009/4/14 閲覧）.

**Cluster Analysis by Manual Method for Educational Purpose** : Tomohiko YOSHIDA[1)], Anas[1)] and Shun-ichi KOBAYASHI[2)] ([1)]*Fac. of Agr., Utsunomiya Univ., Utsunomiya 321-8505, Japan;* [2)] *Tochigi Agr. Exp. Stn.*)

**Abstract** : Cluster analysis is usually performed by using a computer. We tried manual cluster analysis without using a computer for educational purposes. Data used were DNA markers in random amplified polymorphic analysis of barley cultivars. Then number of DNA markers showing a different number of bands between the cultivars (different markers) was used as the distance. The pair with the fewest number of different markers was decided as the first cluster. Next, the difference between the mean number of different markers in the first cluster and the number in the other cultivars was calculated, and the second cluster with the least difference was decided. The same procedure was continued to decide the following clusters. Two-rowed and six-rowed barley cultivars with different origins showed a satisfactory dendrogram. Computer analysis gave the same result. This method can be easily understood and has a good educational effect.

**Key words** : Cluster analysis, Dendrogram, Education, Manual calculation.