解　説

マハラノビスの距離 入門

－MTS法を理解するために－

An introduction to Mahalanobis distance for MTS methods

救仁郷　誠*
Makoto Qunigoh

1. はじめに

「マハラノビスの距離」とは，多変量 (多次元) 空間における距離尺度[1)2)a]のひとつです。

従来，多変量解析の判別分析で用いられているもので[b]，その道のテキストにはたいてい載っています。ただ，一般にはそれほどポピュラーではありません。

品質工学のなかで，この「マハラノビスの距離」を使っているのが，ここ何年か話題になっているMTS法です。MTS法がどういうものかを簡単に言うのは難しいことですが，勇気を奮って言えば『物事の状態の正常さ加減を量る方法論』だと思います。対象となるのは，ひとつやふたつの特性値では表せない状態，例えば健康や天候など，また衣服の着心地などといったものも含まれています。「マハラノビスの距離」は，それらに対する『ものさし』として，まさにMTS法の主役です。

さて，このMTS法については，田口先生をはじめとして多くの方々から，あまたの解説・実施例などがあります。ですから，いまさら何を，といった感はあります。しかし，一方では，いまだによくわからない，また新人に説明しやすい読み物が欲しいといった声も，まだまだあるようです。

そういうわけで，屋上に屋を重ねることにはなりますが，入門者のみなさんのために，MTS法の基本ともいうべき「マハラノビスの距離」がどういうもので，何を意味しているのかを主に述べていきます。

*富士ゼロックス株式会社，正会員10009

2. MTS法って何？

MTSとは，Maharanobis Taguchi System の略[c]です。

最初のMaharanobisは人名です。もう故人になられていますが，インドの著名な統計学者さんで，前述の「マハラノビスの距離」を考え出された方だそうです。次のTaguchiは，品質工学の田口玄一先生です。

ですから，MTS法は，名称から言えば，「マハラノビスの距離」を利用した品質工学の手法のひとつ (System) だということでしょう。ねらいは，予測や診断のためのパターン認識[3)]で，具体的には物事の状態の正常さ加減を量ることになります。

マハラノビスの距離の使い方も，多変量解析の場合とは異なり，基準となる多変量データ群からどれだけ離れているかを示す，逆に言えば，どれだけ似ているかをあらわす尺度として用います。

もう少し言えば，定義された基準データ群だけから「原点と単位量」(＝マハラノビス空間)をつくり，それを「ものさし」に使う方法です。

この「基準データ群だけから」という点にも特徴があります。

たとえば，基準を正常状態としましょう。判別分析や重回帰分析などでは，異常の情報も与えなければ尺度がつくれません。ところが異常さが多様であればあるほど，すべて考慮するわけにもいかず，実際のところ限定された尺度になってしまいます。その点，このMTS法では，基準からだけで作りますので，「どのくらい」基準と違うかについては汎用性の高い尺度といえます。ただし，「どのように」違うのかをあらわすものではありませ

ん。

なお,「マハラノビスの距離」という用語の使われ方を見ていますと,「ものさし」を指す場合，またそれで測った「距離の値」の場合,さらには「MTS法」全体をと，いろいろあるようです。

2.1　MTS法のポイント

MTS法の詳細を見ていきますと，以下の5つのポイントがあります。また適用する際は，この順にステップを踏んでいきます。

① 基準となる正常データ群の選び方
② そのデータ群から「マハラノビスの距離」を計算する方法，そしてその意味合い
③ 求めた「マハラノビスの距離」が，尺度として妥当かどうかを評価する方法(SN比)
④ 「マハラノビスの距離」の妥当性を損なわないように，変量の数を減らす方法
⑤ 「マハラノビスの距離」のしきい値の求め方

これらのなかで，重要なのは①ですが，基本となるのは，やはり②の「マハラノビスの距離」自身のことになります。

3. MTS法における「マハラノビスの距離」

冒頭にも書きましたが，MTS法では，基準のデータ群にどれだけ似ているかを示す多変量空間の測度として「マハラノビスの距離」を用いています。

さて，この距離の意味合いについてお話していくわけですが，その前に，多変量ではなく1変量の場合について考えてみましょう。

3.1　1変量の場合の距離

唐突ですが，皆さんがよくご存知の「偏差値」を思い出してみましょう。

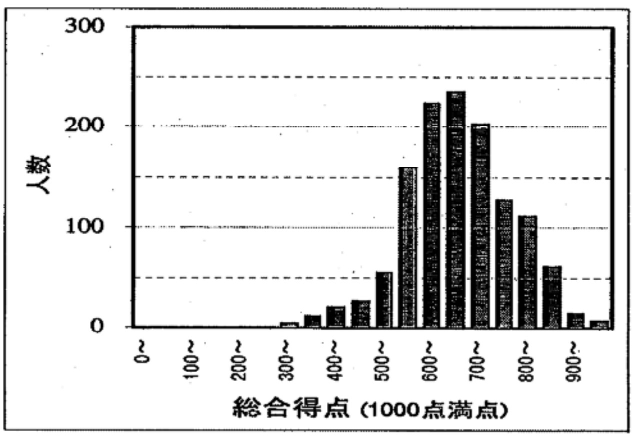

図1　模擬試験の得点分布

図1は，ある模擬試験の得点と人数の分布をあらわしたものです。この試験で785点を取ったA君の偏差値は，59.4でした。

さて，この偏差値というものはなにを意味していたのでしょうか。

3.1.1　偏差値

$$=\frac{x_i-\bar{x}}{\sigma}\times 10+50 \tag{1}$$

$$=\frac{785-680.5}{111.1}\times 10+50\cong 59.4 \tag{1'}$$

(ただし，平均点$\bar{x}$ ＝ 680.5，標準　偏差σ＝111.1)

この偏差値の式は，ちょうど平均点のとき50，平均点＋標準偏差のとき50＋10＝60となるように作られています。

つまり偏差値は，平均と標準偏差で正規化されているわけで，平均点$\bar{x}$を基準点(50)とし，標準偏差σの10分の1を単位量とする尺度になっているわけです。また，特に1式の第1項だけについていえば，平均点を原点とする尺度で，その値は受験者集団の平均からの距離ともいえるわけです。

なにかと悪名の高い偏差値ですが，これもまた立派な「基準からの距離」になっています。

3.2　多変量の場合の距離

そして，多変量です。まずは最もわかりやすい2変量の例を見ていきましょう。

図2aは，数学(x)と物理(y)の100人分の得点データを示した散布図です。

一目でわかりますように，数学と物理の得点には高い相関があります，実際これらの両者の相関係数は0.8でした。また全体に数学よりも物理の得点が低く，回帰係数は0.6でした。これは，数学の問題がやさしかったということなのか，それともそういう受験生たちだけが集まっているのか，ここにあるデータだけではわかりません。

さて，ここでこのような特徴をもった100人の集団に対して「似ているかどうか」ということについて考えます。

基準点は，数学と物理それぞれの平均点の座標で問題ないでしょうが，スケールについてはどうでしょうか。この例では同じ100点満点の点数が単位で

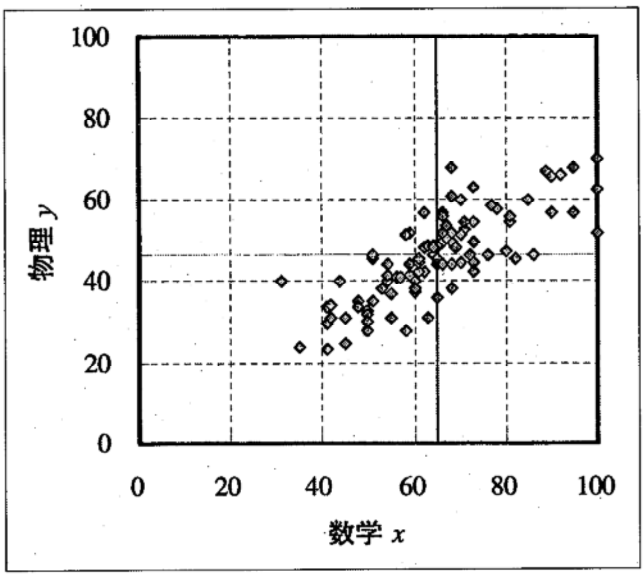

図2a 数学・物理の得点分布

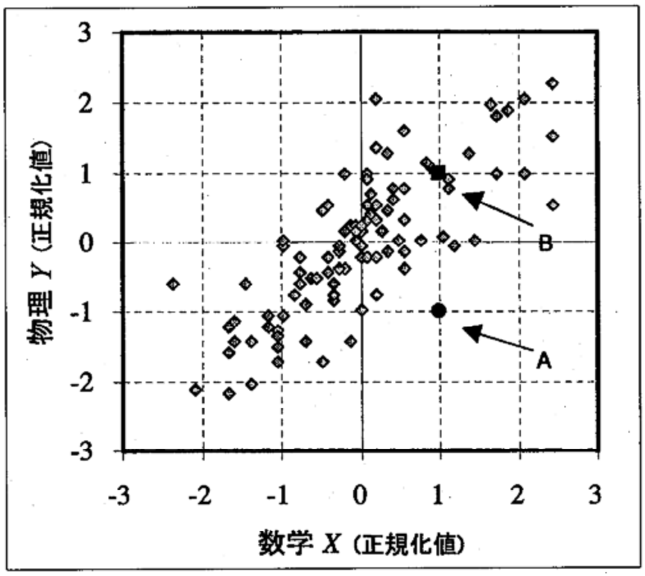

図2b 数学・物理の得点分布 (正規化値)

すので，このままで良いという考え方もあります。一方，分布の範囲を揃えるようにスケールをつけ直すやり方もあります。どちらが適当かはケースバイケースですが，変量間で単位や数値の範囲・桁が異なることが多いので，偏差値と同様に平均と標準偏差で正規化することが一般的です。

$$X_i = \frac{x_i - \overline{x}}{\sigma_x},\quad Y_i = \frac{y_i - \overline{y}}{\sigma_y} \tag{2}$$

そこで，さきほど数学 (x) と物理 (y) の例についても，2式を使って正規化し，プロットしなおしました。

図2bの各軸のスケールは，2式で正規化された値 (X, Y) です。数学と物理が共に平均点なら原点 (0, 0) に，両方が平均＋標準偏差の得点であれば (1, 1) にプロットされます。元々の図2aと比べると，ずいぶん印象が変わり，なにかしら偏ったところというかクセが消え，相関の有り無しだけが目立つ図柄になりました。正規化とはこういうことになります。

さて，別のところで，A君とB君とが同じ試験を受けたとしましょう。それぞれの得点により，二人は図2bのA点 (1,-1) とB点 (1,1) にプロットされました。

ここで，A君とB君とでは，どちらが，100人の集団に近いかを考えてみましょう。成績が良いとか悪いとかではなく，基準となる集団に「似ているかどうか」という観点です。

3.2.1 ユークリッド距離

まずは,集団の中心からの距離を見てみましょう。すでに正規化されたスケールで描かれている図2bでは原点 (0, 0) が中心ですので，そこからA,Bの各点を結ぶ線分の長さdを，おなじみのピタゴラスの定理を使って計算してみます。

ちなみに，普段はいいませんが，こういったごく普通に使っている長さをユークリッドの距離といいます。

$$d^2 = X^2 + Y^2 \tag{3}$$

ただし，この場合に限って，次項で説明するマハラノビスの距離と整合をとるため，この式に係数kを付けた次式を使って計算します。この場合は$k=2$です。

$$d^2 = \frac{1}{k}\left(X^2 + Y^2\right) \tag{3′}$$

$$A\,(1,-1):\quad d_A = \sqrt{\frac{1^2 + (-1)^2}{2}} = 1 \tag{3′a}$$

$$B\,(1,\ 1):\quad d_B = \sqrt{\frac{1^2 + 1^2}{2}} = 1 \tag{3′b}$$

結果は見た目の通りで，A君とB君共に，集団の中心からの距離は1，つまり「ユークリッドの距離」という尺度で見れば，二人には差が無いということになりました。

そうは言っても，「数学と物理の得点に相関がある」という集団の特徴からすれば，数学の点数に比べて物理が低いA君はどうみても変わり者です。それなのにA君とB君が同じというのは腑に落ちません。

ただ，そのように感じても，そもそもユークリッドの距離というのはそういうことを考慮してない尺度です。この集団の相関があるという特徴を観点に入れるのなら，これは不適当な尺度だというしかありません。

3.2.2　マハラノビスの距離

そういったわけで，変量間の相関を考慮していることを特徴とする「マハラノビスの距離」の出番となります。

例のように，2変量で，それぞれが正規化されたデータの場合，「原点から任意の点 $(X,\ Y)$ までのマハラノビスの距離D」は以下の，相関係数行列の逆行列を含む2次形式から求めます。

$$D^2=\frac{1}{k}\begin{bmatrix}X & Y\end{bmatrix}\begin{pmatrix}1 & r\\ r & 1\end{pmatrix}^{-1}\begin{bmatrix}X\\ Y\end{bmatrix} \tag{4}$$

$$=\frac{1}{k}\begin{bmatrix}X & Y\end{bmatrix}\cdot\frac{1}{1-r^2}\begin{pmatrix}1 & -r\\ -r & 1\end{pmatrix}\begin{bmatrix}X\\ Y\end{bmatrix} \tag{4.1}$$

$$=\frac{X^2-2rXY+Y^2}{k\left(1-r^2\right)} \tag{4.2}$$

ここで，$X,\ Y$は数学と物理の正規化した得点，rはそれらの相関係数です。なお，kは変量の数で，この場合は2です。

これに，実際の相関係数r＝0.8と，それぞれA君とB君の値を代入して，

☆ A君: 数学(X=1)、物理(Y=−1)

$$D_A=\sqrt{\frac{1^2-2\times0.8\times1\times(-1)+(-1)^2}{2\left(1-0.8^2\right)}} \tag{4$'$a}$$

$$=\sqrt{\frac{3.6}{0.72}}=\sqrt{5.0}=2.2$$

☆ B君: 数学(X=1)、物理(Y=1)

$$D_B=\sqrt{\frac{1^2-2\times0.8\times1\times1+1^2}{2\left(1-0.8^2\right)}} \tag{4$'$b}$$

$$=\sqrt{\frac{0.4}{0.72}}=0.75$$

このように，マハラノビスの距離では「数学に比べて物理のできが劣るA君は，基準の集団（原点）からのB君の距離よりも約3倍離れたところに位置している」という結果になりました。

先ほどのユークリッドの距離の結果と比べてみて，どう思われるでしょうか。

さて，このマハラノビスの距離Dがどういうものなのか，もう少し見てみましょう。

4.2式を少し変形した次の式を見てください。

$$D^2=\frac{1}{k}\left(\frac{X^2+Y^2-2rXY}{1-r^2}\right) \tag{4.3}$$

まず，相関のない場合，つまり相関係数 $r=0$ のときには，括弧内の分子の第3項が消え，分母が1になってユークリッドの距離（3′式）と全く同じ形になります。

一方，相関がある場合には，分子の第3項によって，ユークリッドの距離よりも近くなったり，遠くなったりします。例えば数学と物理の例のように正の相関 $(r>0)$ がある場合には，第1，3象限側のデータ $(XY>0)$ なら比較的近くに，それと直交する方向の第2,4象限側 $(XY<0)$ では遠くに，観測される傾向にあります。

ですから，マハラノビスの距離は，ユークリッドの距離に，相関とデータの位置関係による効果を足し引きしたものだといえます。

4.3式からいえば以上のようになるのですが，わかりにくいのでもう少し感覚的に見るため，二種類の図を示します。

まず，先ほどの数学と物理の図2b $(r=0.8)$ に，等高線ならぬ「等マハラノビスの距離線」を書き加えたのが図3です。

等距離線の間隔は，マハラノビスの距離Dの値で0.5です。A君は$D=2$の線の外側にいて，B君との間には等距離線3本があり，それだけ隔たっていることがひと目でわかるようになりました。

これらの閉曲線を，原点のまわりに45°回転させた後の座標軸を$(L,\ S)$とすると，4.3式は次のよう

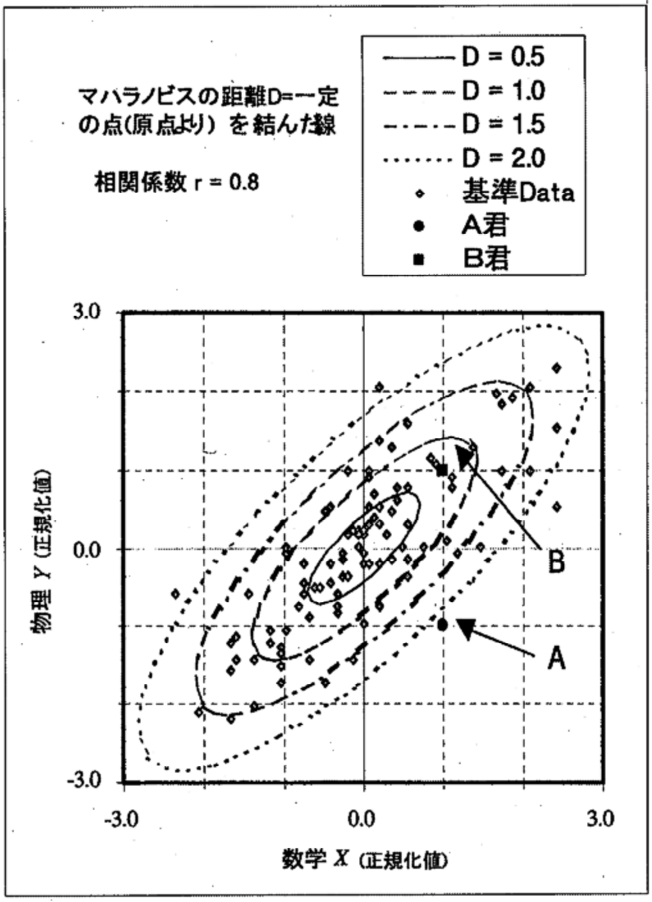

図3　等マハラノビス距離線 (r =0.8)

に書き換えられます。

$$D^2 = \frac{1}{k}\left(\frac{1}{1+r}L^2 + \frac{1}{1-r}S^2\right) \tag{4.4}$$

つまり，「等マハラノビス距離線」は相関係数によって長径・短径の比が規定される楕円です。

つづいて相関係数の影響を示したのが，もうひとつの図4です。

このように，相関のないときには真円ですが，相関係数が高くなるにつれて細りとした楕円になってきます。なお，図の場合は，相関が正の例なので右上がりの楕円ですが，負の場合には右下がりになります。

また，この楕円の径自体はDに比例します。ただし，その比例係数は方向によって変化し，長径側で最大，短径で最小となります。

以上，2変量の例で，マハラノビスの距離の方が「相関のある集団に似ているかどうか」をあらわすのに適しているということと，その特徴を述べました。

では，3変量以上の場合でもそうなのか，という

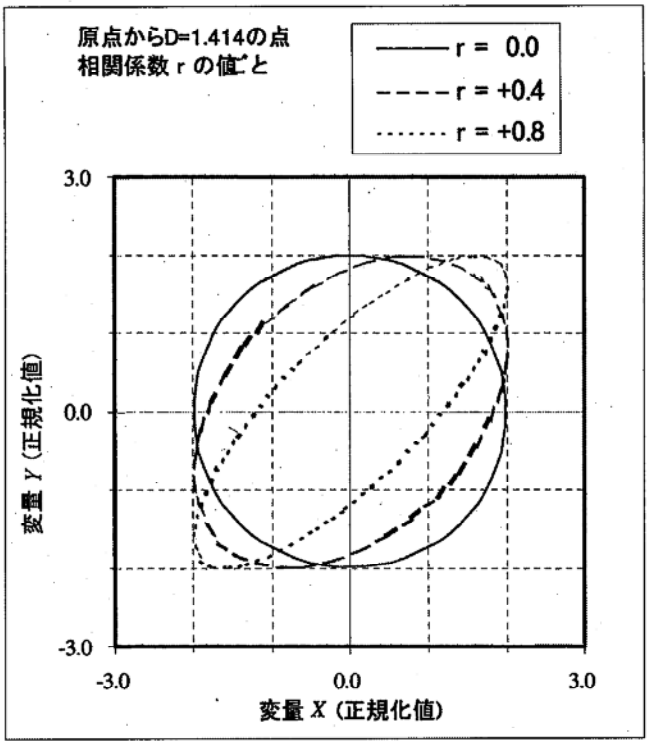

図4　相関係数とマハラノビスの距離

ことになりますが，大雑把にいって2変量の場合の延長線上にあるものとご理解ください。なぜなら，マハラノビスの距離は変量がいくつであっても，2変量間の相関を考慮して求める距離だからです。詳しくは次章をご覧ください。

4. マハラノビスの距離の計算方法

さて，何度も繰り返してしまいますが，MTS法における「マハラノビスの距離」は，基準としたデータ群にどれだけ似ているかを示す尺度です。

ですから，最も重要なものは「基準となるデータ群」で，それらが基準とするのに相応しいだけのデータかどうかが肝心です。具体的には，データ数や欠測値のことだけでなく，基準のなかでの偏りや抜け，重要な変量のとりこぼしを含みます。

これらのことは，結果や応用性が大きく左右しますので，充分に吟味してください。

4.1 計算手順

では，「基準データ群」には問題がない，という前提で，一般的な計算の手順[4]について説明します。

4.1.1 「基準データ群」の用意

「基準となるデータ群」を，下記のように，k個

の変量を持つ，n組のデータの形で用意しておきます。

$$
\begin{array}{c|cccc}
 & \multicolumn{4}{c}{\text{変　量}} \\
 & y_1 & y_2 & \cdots & y_k \\
\hline
1 & y_{11} & y_{12} & \cdots & y_{1k} \\
2 & y_{12} & y_{22} & \cdots & y_{2k} \\
\vdots & \vdots & \vdots & & \vdots \\
n & y_{n1} & y_{n2} & \cdots & y_{nk} \\
\hline
\overline{y} & \overline{y}_1 & \overline{y}_2 & \cdots & \overline{y}_k \\
\sigma & \sigma_1 & \sigma_2 & \cdots & \sigma_k
\end{array}
\qquad (5.1)
$$

データ数nがどれくらいあればよいかというと，多変量解析のテキスト[5)]には100以上と，品質工学の方には「少なくとも数十以上，できれば数百以上が望ましい」[6)]とあります。また，これは個人的な記憶ですが，何年か前に「計量値なら最低200，計数値なら1000」と，うかがったこともありました。

変量の数については，特に言われていることなさそうです。

4.1.2 「基準データ群」の正規化

変量ごとに,それらの平均と標準偏差をもちいて,それぞれのデータを正規化しておきます。

$$
\Rightarrow\left(Y_{ij}=\frac{y_{ij}-\overline{y}_j}{\sigma_j}\right)\Rightarrow
\begin{array}{c|cccc}
 & \multicolumn{4}{c}{\text{変　量}} \\
 & Y_1 & Y_2 & \cdots & Y_k \\
\hline
1 & Y_{11} & Y_{12} & \cdots & Y_{1k} \\
2 & Y_{12} & Y_{22} & \cdots & Y_{2k} \\
\vdots & \vdots & \vdots & & \vdots \\
n & Y_{n1} & Y_{n2} & \cdots & Y_{nk} \\
\hline
\overline{Y} & 0 & 0 & \cdots & 0 \\
\sigma & 1 & 1 & \cdots & 1
\end{array}
\qquad (5.2)
$$

4.1.3 変量間の相関係数行列の算出

k個ある変量から，2個を取り出すすべての組み合わせ（${}_kC_2$通り）について，相関係数を求めておきます。

すでに5.2式で正規化されたデータの場合は，次式で相関係数が求められます。

$$
r_{pq}=r_{qp}=\frac{\sum_{i=1}^{n}\left(Y_{ip}\cdot Y_{iq}\right)}{\sqrt{\sum_{i=1}^{n}Y_{ip}{}^2\cdot\sum_{i=1}^{n}Y_{iq}{}^2}} \qquad (5.3)
$$

対角要素を1，その他のp行q列の要素を上のrpqとしたk×kの相関係数行列Rを作ります。

$$
R=\begin{pmatrix}
1 & r_{12} & r_{13} & \cdots & r_{1k} \\
r_{21} & 1 & r_{23} & \cdots & r_{2k} \\
r_{31} & r_{32} & 1 & \cdots & r_{3k} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
r_{k1} & r_{k2} & r_{k3} & \cdots & 1
\end{pmatrix} \qquad (5.4)
$$

なお，この行列は基準データ群の，いわば標準偏差のような意味合いをもったものです。多変量データは，1変量が平均と分散（標準偏差の二乗）とで要約されるように，平均ベクトルと分散共分散行列で要約されます。ここでは，正規化データが対象ですので，各変量の分散＝1，共分散＝相関係数となり，分散共分散行列と相関係数行列は全く同じものになります。

4.1.4 相関係数行列の逆行列の算出

上の相関係数行列Rの逆行列Aを求めます。これも対称行列になります。

$$
A=\begin{pmatrix}
a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\
a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\
a_{31} & a_{32} & a_{33} & \cdots & a_{3k} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kk}
\end{pmatrix}=R^{-1} \qquad (5.5)
$$

4.1.5 マハラノビスの距離の算出

任意のデータ(1組）と上の逆行列の二次形式により，そのデータの「マハラノビスの距離D」を求めます。

$$
D^2=\frac{1}{k}\begin{pmatrix}Y_1 & \cdots & Y_k\end{pmatrix}\begin{pmatrix}
a_{11} & \cdots & a_{1k} \\
\vdots & \ddots & \vdots \\
a_{k1} & \cdots & a_{kk}
\end{pmatrix}\begin{pmatrix}Y_1 \\ \vdots \\ Y_k\end{pmatrix} \qquad (5.4)
$$

$$
=\frac{1}{k}\sum_{p=1}^{k}\sum_{q=1}^{k}a_{pq}Y_pY_q \qquad (5.5)
$$

なお，MTS法では，上記の5.4, 5.5式のように，変量の数kで割った値をD^2としています。これは基準データ群の「マハラノビスの距離D」の二乗平均を，変量の数にかかわらず1前後に調整するための処置で，後述の「変量を減らす方法」の時に意味をもってきます。

4.2 マハラノビスの距離を計算する際の問題点

計算自体は，式に代入するだけですので，いたって単純です。逆行列を求めるのに多少の手間がかかりますが，やり方は行列または数値計算のテキスト[7]に載っています。また最近では計算ツールの類[d]が出回っていますので，それらを利用するのが，早くて便利で，あまり間違いがありません。

ただし，ツールを使う，使わないにかかわらず，途中で計算ができなくなる，もしくはさじ加減次第で結果が大きく変わってしまうケースがあります。そこで，その際の対処の方法を簡単に列挙します。

4.2.1 変量の値を正規化できない。

変量の値がすべて同じ場合には，標準偏差がゼロになってしまいますから，正規化するときに割り算しようとしてもできないわけです。対処方法としては，次の2通り。

- その変量を計算に使わない。
- 標準偏差に仮の値を使う。それができない計算ツールの場合には，データの1～2個の値をちょっとだけ変えて標準偏差を発生させる。

基準のデータ群と似ているかどうかを測るために意味のない変量でしたら，前者の方法を。その値がちょっとでも違っていたら，それだけで違うぞ，ということでしたら後者の方法を選んでください。ただ標準偏差の影響は大きいと思います，いくつかの値で試してみたほうが良さそうです。

4.2.2 逆行列が計算できない。

データ数が変量数以下の場合は，逆行列が存在しません。この場合はデータを増やす[e]か，変数を減らすしかありません。同数であっても正規化しているため同様になります。

変量間の相関係数がほとんど1という場合にも，逆行列を求められません。これは，2つの変量が同じものだということですから，どちらかを計算から除外します。

なお，計算ツールによっては，逆に，本来計算できないはずのケースでも，桁落ち誤差などにより，逆行列を出力してしまうことがあります。なかなか見つけにくいとは思いますが，おかしいと感じたら，元の行列との掛け算で単位行列になるかどうか試してみてください。

4.2.3 欠損データの扱い

データの一部に，測定できなかった，あるいは記録がなくなってしまったなどの理由で，値がない場合には，

- 仮の値を入れる。
- 値のないデータを計算から除外する。

仮の値としては，基準データ群の変量ごとの平均値を使うのが一般的です。意図的にではなく他の値を入れてしまうと，結果が大きく違ってしまうことがあるので要注意です[8]。

いずれにせよ，これらはデータがないときにやむを得ず取る方法です。これからデータを採る場合には，欠測のないようにしてください。

4.3 他の計算方法

前記のように，相関係数行列の逆行列をまともに解いていくやり方の他に，基準データ群で定義される多次元空間の方を先に直交化しておいて，そこから距離を求めるやり方があります。簡単に言えば，前章に出てきた「等マハラノビス距離線」を，楕円でなく真円になるように軸を取り直してから距離を測るという方法です。

直交化の方法には，シュミット(Schmidt)の直交展開[9]をはじめとして，主成分[10]やそれをさらに回転させる方法，また特異値分解[11]を利用するやり方などがあります。

また，直交化した後の相関係数行列は単位行列になりますので，逆行列の計算が不要でその手間が省けます。ただ，変量は，もともとの変量をいくつか結合したものになり，直交化のやり方によっていろいろな特徴がでてきます。詳細につきましては，それぞれの文献を参照してください。

なお，どの方法にしましても，マハラノビスの距離の計算過程に差異があるだけです。同じ基準データ群で，同じだけの変量を使えば，まともに計算した場合と同じ値になります。

5. その他のポイント

ここまでMTS法における「マハラノビスの距離」がどういうもので，どう計算するかを述べてきました。

ここからは残りの，③マハラノビスの距離の評価方法と，④変量の数を減らす方法について，簡単に紹介します。ただし，⑤しきい値の求め方について

は，損失関数を使ってという程度にとどめます。

詳しくは，成書を参照してください。

5.1 「マハラノビスの距離」を評価する方法。

基準データ群から作った「マハラノビスの距離」という尺度が妥当であるかどうかを，SN比を使って評価する方法です。

以下の，2つの方法が提案されています。

ひとつは，真値といっても良さそうな評価値がある場合[f]で，その値Mとマハラノビスの距離Dとの理想関係を「比例すること」，つまり次式のような動特性と定義した場合の方法です。

$$D = \beta M \tag{6}$$

「評価値M」の異なるデータを複数取りあげ，それと，求めた「マハラノビスの距離D」からゼロ点比例式のSN比を計算し，これを評価尺度とします。

もうひとつは，基準外データの「マハラノビスの距離D」が大きい値になればなるほど，すなわち感度の高いことが良いこととする考え方です。この場合には，基準外データの「マハラノビスの距離D」から求めた望大特性のSN比を使います。

どちらの方法にしましても，SN比を求めるためには基準外のデータが必要になります。この際に，どういうデータを用いるのが良いのか，たとえば異常の多様性について配慮するのかなど，まだ解釈できない点があります。今後の進展を待ちたいと思っています。

5.2 必要な変量の数を減らす方法

さて，この「マハラノビスの距離」ですが，もともと多変量の距離でもあるため，この値を算出するためには「多くの変量」のデータを使います。しかし，実際の評価や検査の場面では，変量が多ければそれだけ測定項目が増え，一般に工数・コストがかかるということになります。

そこで，尺度としての妥当性を損なわないように，また効率的に，計算に用いる変量の数を減らす方法が提案されています。

2水準系の直交表を用意し，その列に変量を，各々第1水準のときには計算に使用，第2水準では使わないというようにわりつけます。そしてその直交表の行ごとに「マハラノビス空間」，具体的には相関係数行列の逆行列を求め直し，さらに前節の方法でSN比を求めていきます。

すべての行のSN比が求まったところで，要因効果を解析し，あってもなくてもいいような変数から順次減らしていくという方法[12]です。

それから，もうひとつには，シュミット(Schmidt)の直交展開[13]を利用した方法もあります。

6. まとめ

以上，主にMTS法におけるマハラノビスの距離についてのべてきました。

最後に，特徴を列挙します。

① 基準データ群にどれだけ似ているかを，変量間の相関も考慮して表す，尺度である
(どのように違うのかを示すものではない)

また，相関がどのように考慮されているかというと，基準データ群の中心からの距離は，

② 相関のある方向では近くに，そうでない方向では遠くにと表される

③ 上の傾向は，相関係数の絶対値が大きくなるほど顕著になる

最後に，この文がMTS法を勉強または利用される方々の参考になれば幸いです。

参考文献

1) B.F.J.マンリー著，村上正康・田栗正章 共訳 「多変量解析の基礎」 培風館 (1992)，p.37
2) 加藤，安倍，根元：改良型マハラノビス距離を用いた高精度な手書き文字認識，信学論(D-II)，Vol.J79-D-II，No.1 (1996)，pp.45-52
3) 田口玄一「品質工学の数理」日本規格協会 (1999)，p.134
4) 文献3，p.136
5) 文献1，p.43
6) 文献3，p.137
7) 近藤，高橋，他「統計・数値解析」培風館 (1968); 戸川隼人「BASIC数値計算プログラム集」サイエンス社 (1986) など
8) 長谷川，他　「マハラノビスの距離を用いた健康診断と欠測値対策　パート1 欠測値がない場合」　品質工学 Vol.5，No.5 (1997) ; 同 「パート2 欠測値を放置して自動的にゼロが代入されたケースについて」　品質工学 Vol.5，No.6 (1997) ; 同 「パート3 欠測値対策を行なったケースについて」 品質工

学 Vol.6, No.3 (1998)

9) 文献3, p.147；長谷川, 他 「シュミットの直交展開を用いた健康診断の信頼性向上と経済性評価」第7回品質工学研究発表大会論文集 p.81 (1999)

10) 中津川, 山本, 大内：MTSにおけるマハラノビス距離, 品質工学Vol.8, No.5, p.19 (2000)

11) 松田, 池田, 鴨下, 東原：マハラノビス空間直交化の故障診断への適用, 品質工学Vol.8, No.2, p.61 (2000)

12) 長谷川, 他 「MTSを用いた健康診断による総合判定の信頼性向上」 品質工学 Vol.7, No.2 (1999)；金沢, 他 「MTSによる処理液診断システムの構築」 品質工学 Vol.6, No.6 (1998)；手島, 他 「マハラノビス・タグチ・システム法を適用した外観検査技術の研究」 品質工学 Vol.5, No.5 (1997)

13) 田口玄一「品質工学の数理」日本規格協会 (1999), p.147；長谷川, 他 「シュミットの直交展開を用いた健康診断の信頼性向上と経済性評価」 第7回品質工学研究発表大会論文集 p.81 (1999)

脚注

a 他には, ユークリッド距離, シテイブロック距離等々, 尺度の種類はかなりの数にのぼります。それらのなかあって, 変量間の相関を考慮していることが「マハラノビスの距離」の大きな特徴になっています。

b 多変量解析のひとつ, 判別分析のところに出てきます。「マハラノビスの汎距離」と表記されているものもあります。

c ということに落ち着きつつあるようです。

d 田口先生のお宅でMaharanobis博士の写真を見せていただいたことがあります。いっしょに写っているのが, お若い頃の田口先生, さらにネール首相, シューハート博士。

e 変数選択や項目選択という呼び方がされています。

f 市販のものでは, オーケンのMTS, それと日本規格協会のANOVA, ASIの・・・などがあります。

g ただし, 同じデータの繰り返しで増やしても, 相関係数行列のランクはそのままですので効果がありません。

h 真値と言ってもいいような値があるなら, わざわざ「マハラノビスの距離」など求めることはない, というご意見もあるかもしれませんが, どちらを採るかはコストと精度のバランスで選ぶものかと思います。

発表論文の学会誌への投稿のお願い

研究発表大会で発表された論文を討論も踏まえ, 学会誌に投稿して下さい。口頭発表だけでなく, 誌上発表をすることにより研究として認められることになります。大会での討論をQ&Aとして, 論文の最後に入れて下さい。

特に, 雑誌などに紹介される可能性のあるときは, 至急投稿しないと既発表ということで論文投稿の権利を失います。

(編集委員会)