

# MTS 法と各距離における分析法の比較

棚橋 誠 松田 眞一

タグチメソッドには正例事象群に基づくマハラノビス距離の算出により検査事象を正例と負例に分類する MTS 法という方法がある。本論文では変数選択には直交表および要因効果図を適用し、閾値決定には正例と負例の誤判別率の平均を適用することによって MTS 法の全自動化を R 上で実現させた。また、MTS 法以外の各距離における分析法も自動化し比較を行なった。シミュレーションを実行した結果、各距離における分析法の有効性を確認することができた。

## 1 はじめに

品質問題発生背景は企業によって様々であるが、顧客を相手にしている以上、企業にとって品質とは最も重要なことであり、欠くことのできない項目であると考えられる。問題が発生してしまったら、早急に解決策を考案し、それを実行すべきである。しかし、それよりも問題発生を未然に防ぐことのほうが重要である。問題発生を未然に防ぐ方法はいくつか挙げられるが、実験計画法を基に 1950 年代から田口玄一博士によって考案されて、問題発生のカギとなる製品のばらつきや劣化をなくす方法として発展したタグチメソッドもその 1 つである。なお、実験計画法が平均値の変化の解析であるのに対して、タグチメソッドは問題発生を未然に防ぐ予防設計技術である。タグチメソッドには MTS 法 (Mahalanobis Taguchi System 法) という方法が存在する。これは、正常空間である正例事象群に基づき検査事象のマハラノビス距離を算出し、検査事象を正例と異常空間である負例とに分類する方法である。

本論文では、MTS 法を基に MTS 法の距離と吉野ら (2004) で提案されている主座標分析法や対応分析法といった MTS 法以外の距離を用いた分析法の全自動化を R 上で実現させ、そのシミュレーションを行なう。

## 2 MTS 法の概要

MTS 法において、検査事象を分類するには正例事象群と負例事象群が必要とされる。正例事象群からデータの中心を求めて、その中心からの距離によって正例か負例かに分類される。以下にその概要を記す。(田口 (1988), 立林 (2004), 中津川・大内 (2000) 参照)

### 2.1 MTS 法の距離

表 1 に示される  $n$  事象  $k$  変数の多変量データを正例事象群とする。また、表 2 に示される  $m$  事象  $k$  変数の多変量データを負例事象群とする。

正例事象群データ  $x_{ij} (i = 1, 2, \dots, n; j = 1, 2, \dots, k)$  に基づき、各変数の平均  $\bar{x}_j$  及び標準偏差  $s_j$  を求める。

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

表 1 正例事象群データ

| 事象番号     | 変数       |          |          |          |
|----------|----------|----------|----------|----------|
|          | $x_1$    | $x_2$    | $\cdots$ | $x_k$    |
| 1        | $x_{11}$ | $x_{12}$ | $\cdots$ | $x_{1k}$ |
| 2        | $x_{21}$ | $x_{22}$ | $\cdots$ | $x_{2k}$ |
| $\vdots$ | $\vdots$ | $\vdots$ |          | $\vdots$ |
| $n$      | $x_{n1}$ | $x_{n2}$ | $\cdots$ | $x_{nk}$ |

表 2 負例事象群データ

| 事象番号     | 変数       |          |          |          |
|----------|----------|----------|----------|----------|
|          | $y_1$    | $y_2$    | $\cdots$ | $y_k$    |
| 1        | $y_{11}$ | $y_{12}$ | $\cdots$ | $y_{1k}$ |
| 2        | $y_{21}$ | $y_{22}$ | $\cdots$ | $y_{2k}$ |
| $\vdots$ | $\vdots$ | $\vdots$ |          | $\vdots$ |
| $m$      | $y_{m1}$ | $y_{m2}$ | $\cdots$ | $y_{mk}$ |

$$s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$$

正例事象群における平均  $\bar{x}_j$  と標準偏差  $s_j$  を用いて,  $x_{ij}$  と  $y_{i'j}$  の規準化を行なう。

$$u_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

$$v_{i'j} = \frac{y_{i'j} - \bar{y}_j}{s_j}$$

規準化データ  $u_{ij}$  を用いて, 正例事象群の相関係数行列  $R$  を求める。

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{pmatrix}$$

相関係数行列  $R$  の逆行列を  $R^{-1}$  とするとき, 正例における  $i$  番目のデータの MTS 法の距離  $d_{x_i}^2 (i = 1, 2, \dots, n)$  および負例における  $i'$  番目のデータの MTS 法の距離  $d_{y_{i'}}^2 (i' = 1, 2, \dots, m)$  は以下のように求められる。

$$d_{x_i}^2 = \frac{\mathbf{u}_i^t \mathbf{R}^{-1} \mathbf{u}_i}{k} \quad \text{ただし, } \mathbf{u}_i = [u_{i1} \cdots u_{ik}]$$

$$d_{y_{i'}}^2 = \frac{\mathbf{v}_{i'}^t \mathbf{R}^{-1} \mathbf{v}_{i'}}{k} \quad \text{ただし, } \mathbf{v}_{i'} = [v_{i'1} \cdots v_{i'k}]$$

## 2.2 望大特性の SN 比

負例の MTS 法の距離  $d_{y_{i'}}^2 (i' = 1, 2, \dots, m)$  に基づいて, 望大特性の SN 比  $\mu$  を求める。

$$\mu = -10 \log \frac{1}{m} \left( \frac{1}{d_{y_1}^2} + \cdots + \frac{1}{d_{y_m}^2} \right)$$

## 2.3 水準選択

正例事象群の各変数を直交表の第1列から順に割付けて、それぞれの負例のMTS法の距離およびSN比を求める。そして、直交表に割付けた制御因子ごとにSN比の水準平均を求める。求めたSN比の水準平均をもとに要因効果図を作成し、SN比が高くなる水準を選択する。

## 2.4 閾値決定法と誤判別率

閾値は正例及び負例のMTS法の距離を用いて決定する。正例及び負例の誤判別率は以下の式で求める。(田中・脇本(1983)参照)

正例事象群の中で負例と判別された個数を  $a$  としたときの正例の誤判別率  $P_1$  は、

$$P_1 = \frac{a}{n}$$

また、負例事象群の中で正例と判別された個数を  $b$  としたときの負例の誤判別率  $P_2$  は、

$$P_2 = \frac{b}{m}$$

である。これら2つの誤判別率より、

$$\frac{P_1 + P_2}{2} \rightarrow \min$$

となるようなMTS法の距離の値を閾値とする。

## 3 MTS法の距離以外の距離

### 3.1 主座標分析法の距離

正例事象群データ  $x_{ij}$  に基づく行和  $x_{i.}$ 、列和  $x_{.j}$ 、総和  $x_{..}$  および負例事象群データ  $y_{i'j}$  に基づく行和  $y_{i'.$  を用いて、 $x_{ij}$  と  $y_{i'j}$  の2重中心化変数変換を行なう。

$$z_{xij} = x_{ij} - \frac{x_{i.}}{k} - \frac{x_{.j}}{n} + \frac{x_{..}}{nk}$$

$$z_{y_{i'j}} = y_{i'j} - \frac{y_{i'..}}{k} - \frac{x_{.j}}{n} + \frac{x_{..}}{nk}$$

これらを用いて、MTS法の距離と同様に主座標分析法の距離を求める。ただし、逆行列を求める際にはムーア・ペンローズの一般化逆行列を用いる。また、得られた各距離は正例の期待値が1になるように変換する。主座標分析法の距離は、吉野ら(2004)によって提案されている分散共分散行列を用いる方法「主座標分析法の距離①」と相関係数行列を用いる方法「主座標分析法の距離②」の2パターンを適用する。

### 3.2 対応分析法の距離

正例事象群データ  $x_{ij}$  に基づく列平均  $\bar{x}_j$  , 列平均の和  $\bar{x}$  , および行和  $x_{i\cdot}$  を用いたとき, 正例における  $i$  番目のデータの対応分析法の距離  $D_{xi}^2$  は以下のような方法で求められる。

$$D_{xi}^2 = \sum_{j=1}^k \frac{1}{\bar{x}_j} \left\{ \frac{\bar{x}_j}{\bar{x}} - \frac{x_{ij}}{x_{i\cdot}} \right\}^2$$

負例も正例を基準に以下のように求める。

$$D_{yi'}^2 = \sum_{j=1}^k \frac{1}{\bar{x}_j} \left\{ \frac{\bar{x}_j}{\bar{x}} - \frac{y_{i'j}}{y_{i'\cdot}} \right\}^2$$

ただし, 得られた各距離は正例の期待値が1になるように変換する。対応分析法の距離は, 吉野ら (2004) によって提案されている無変換の方法「対応分析法の距離①」と偏差値変換を用いる方法「対応分析法の距離②」と重心化を用いる方法「対応分析法の距離③」の3パターンを適用する。

## 4 統計解析ソフト R におけるプログラム概要

プログラムは, どの距離においても以下のような概要で作成した。

1. 入力されたデータを L12 直交表の第 1 列から順に割付ける。このとき, 第 1 水準を「その変数を判定に使用する」, 第 2 水準を「その変数を判定に使用しない」とする。
2. 判定に使用する変数を用いて, 負例の距離を算出し, 負例の距離を基に望大特性の SN 比を算出する。
3. SN 比の水準平均を算出する。
4. 要因効果図を作成し, 効果の大きさにより変数選択をする。
5. 選択された変数を効果が大きい順に並べ替える。
6. 効果が大きい変数から順に 1 変数ずつ加えながら, 負例の距離を算出し, 負例の距離を基に望大特性の SN 比を算出する。
7. SN 比が最大となるような変数の組み合わせを最適変数とする。
8. 最適化された変数により, 正例と負例それぞれの距離を算出する。
9. 正例の誤判別率と負例の誤判別率の平均が最小となるような閾値を決定する。
10. 正例と負例それぞれの誤判別サンプルを検索する。

## 5 検証実験

以下で 4 つの実データに対して, それぞれの分析法を適用した結果を述べる。(表内の\*は誤判別率の平均が最小のものを示す。)

## 5.1 うつ病データと実験結果

株式会社デンソーの吉野氏に提供していただいたうつ病診断の10次元データである。うつ病と診断されていない人を正例事象群データ(755サンプル)とし、うつ病と診断された人を負例事象群データ(30サンプル)とする。プログラムによる検証実験結果を表3に示す。

表3 うつ病データの実験結果

|        | 最適選択変数             | 閾値   | 正例誤判別率 | 負例誤判別率 | 誤判別率平均 |
|--------|--------------------|------|--------|--------|--------|
| MTS法   | 1,2,3,4,5,6,7,9    | 1.91 | 0.070  | 0.167  | 0.118* |
| 主座標分析① | 1,2,3,4,5,6,7,9,10 | 1.62 | 0.135  | 0.133  | 0.134  |
| 主座標分析② | 1,2,3,4,5,6,7,9,10 | 1.63 | 0.132  | 0.133  | 0.133  |
| 対応分析①  | 1,2,3,4,5,6,7,9    | 1.52 | 0.150  | 0.200  | 0.175  |
| 対応分析②  | 2,3,4,6,7,9,10     | 1.54 | 0.166  | 0.233  | 0.199  |
| 対応分析③  | 1,2,3,4,5,6,7,9,10 | 1.79 | 0.114  | 0.200  | 0.157  |

うつ病データでは、誤判別率の平均がMTS法の距離で0.118と最小の値となった。

## 5.2 株式上場企業データと実験結果

東京証券取引所(2005)から検索した建設業における上場企業の2005年度決算の8次元データである。東証1部上場企業を正例事象群データ(102サンプル)とし、東証2部上場企業を負例事象群データ(26サンプル)とする。プログラムによる検証実験結果を表4に示す。

表4 株式上場企業データの実験結果

|        | 最適選択変数  | 閾値   | 正例誤判別率 | 負例誤判別率 | 誤判別率平均 |
|--------|---------|------|--------|--------|--------|
| MTS法   | 8       | 0.41 | 0.333  | 0.115  | 0.224* |
| 主座標分析① | 1,6     | 0.22 | 0.461  | 0.038  | 0.250  |
| 主座標分析② | 1,6     | 0.22 | 0.461  | 0.038  | 0.250  |
| 対応分析①  | 2,5,6,7 | 0.62 | 0.422  | 0.269  | 0.345  |
| 対応分析②  | 3,4,8   | 3.60 | 0.020  | 0.962  | 0.491  |
| 対応分析③  | 1,6     | 0.48 | 0.412  | 0.077  | 0.244  |

株式上場企業データでは、誤判別率の平均がMTS法の距離で0.224と最小の値になった。

## 5.3 事故分類別交通事故データと実験結果

交通安全マップ(警察庁・国土交通省(2004))から作成された事故分類別交通事故の10次元データである。東名高速道路および名神自動車道、または国道1号線を通らない都道府県を正例事象群データ(37サンプル)とし、東名高速道路および名神自動車道、または国道1号線を通る都道府県を負例データ(10サンプル)とする。プログラムによる検証実験結果を表5に示す。

表 5 事故分類別交通事故データの実験結果

|        | 最適選択変数        | 閾値   | 正例誤判別率 | 負例誤判別率 | 誤判別率平均 |
|--------|---------------|------|--------|--------|--------|
| MTS 法  | 1,2,3,4,7,8   | 2.06 | 0.081  | 0.200  | 0.141* |
| 主座標分析① | 1,2,3,4,5,7,8 | 2.46 | 0.081  | 0.300  | 0.191  |
| 主座標分析② | 1,2,3,4,5,7,8 | 2.46 | 0.081  | 0.300  | 0.191  |
| 対応分析①  | 3,4,5,7       | 1.55 | 0.135  | 0.500  | 0.318  |
| 対応分析②  | 3,4,7,10      | 0.74 | 0.189  | 0.200  | 0.195  |
| 対応分析③  | 7,9           | 1.09 | 0.162  | 0.300  | 0.231  |

事故分類別交通事故データでは、誤判別率の平均が MTS 法の距離で 0.141 と最小の値になった。しかし、このデータの場合は単位系が異なっているため、単位系を統一する、すなわち正例事象群データを基に標準化を行なったからの主座標分析の距離による実験も行った。その結果、主座標分析法の距離は MTS 法の距離と同等の機能を示した。

#### 5.4 気象データと実験結果

気象庁 (2006) から検索した 2006 年 12 月 1 日から 2006 年 12 月 3 日までの全国の観測地別気象の 10 次元データである。人口 100 万人未満の都市を正例事象群データ (423 サンプル) とし、人口 100 万人以上の都市を負例事象群データ (30 サンプル) とする。プログラムによる検証実験結果を表 6 に示す。

表 6 気象データの実験結果

|        | 最適選択変数         | 閾値   | 正例誤判別率 | 負例誤判別率 | 誤判別率平均 |
|--------|----------------|------|--------|--------|--------|
| MTS 法  | 2,3,4,6,7,9,10 | 0.31 | 0.924  | 0.033  | 0.479  |
| 主座標分析① | 3,4,5,7,8,10   | 0.15 | 0.976  | 0.000  | 0.488  |
| 主座標分析② | 3,4,5,7,8,10   | 0.15 | 0.976  | 0.000  | 0.488  |
| 対応分析①  | 6,7,10         | 0.07 | 0.931  | 0.000  | 0.466* |
| 対応分析②  | 4,6,7,9        | 0.05 | 0.976  | 0.000  | 0.488  |
| 対応分析③  | 3,4,6,7,9,10   | 0.83 | 0.421  | 0.533  | 0.477  |

気象データでは、誤判別率の平均が対応分析法の距離①で 0.466 と最小の値になった。

## 6 主成分プロットにおける比較

4 つのデータにおける第 1 第 2 主成分プロットと第 2 第 3 主成分プロットをそれぞれ図 1, 図 2 に示す。ただし、各プロット図において、● は正例データ、× は負例データを示す。

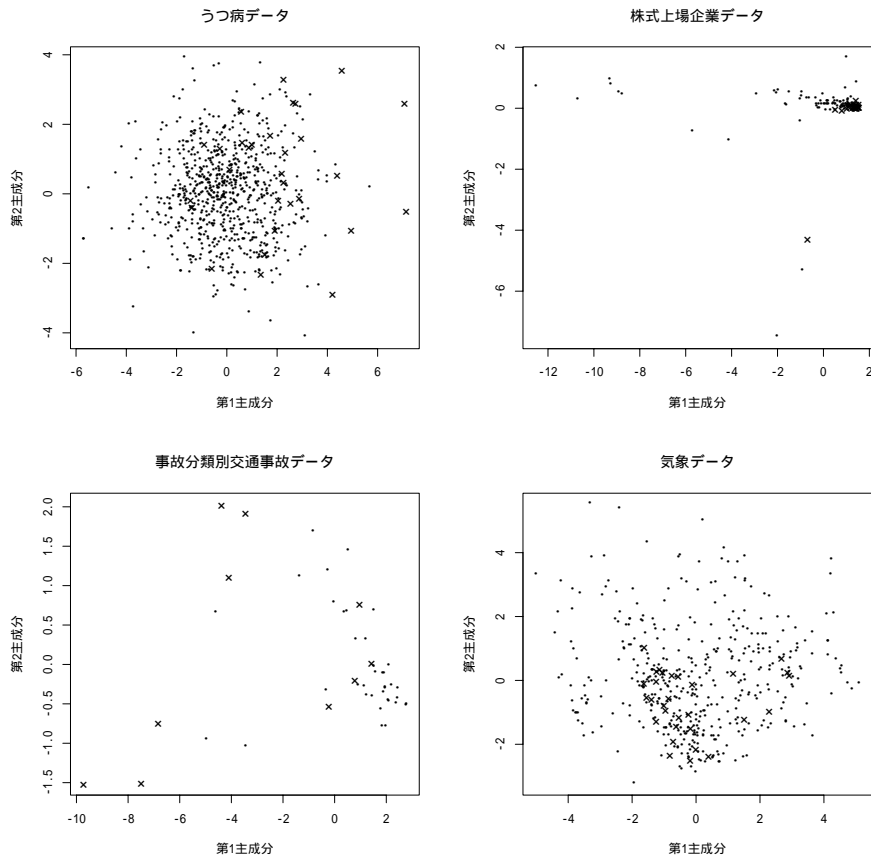


図 1 第 1 第 2 主成分プロット

第 1 主成分において、MTS 法の距離に分類されたうつ病データと株式会社企業データと事故分類別交通事故データは、正例と負例がそれぞれ片寄っているが、MTS 法の距離以外の距離に分類された気象データは、正例と負例に大きな片寄りが見られなかった。

## 7 シミュレーション

前節では、主成分プロットにおける主成分のずれから、データによって有効となる距離に違いが見られた。そこで本節では、主成分のずれを 3 通りの大きさで発生させて、データの違いによる各距離の有効性をシミュレーションする。

### 7.1 プログラム概要

1.  $p$  行 8 列の正例事象群データを発生させる。その際、第 1 主成分は平均 0、標準偏差 1 の正規乱数、第 2 主成分は平均 0、標準偏差  $1/3$  の正規乱数、第 3 主成分は平均 0、標準偏差  $1/5$  の正規乱数で発生させ、表 7 のように各主成分が直交する条件で 3 つの

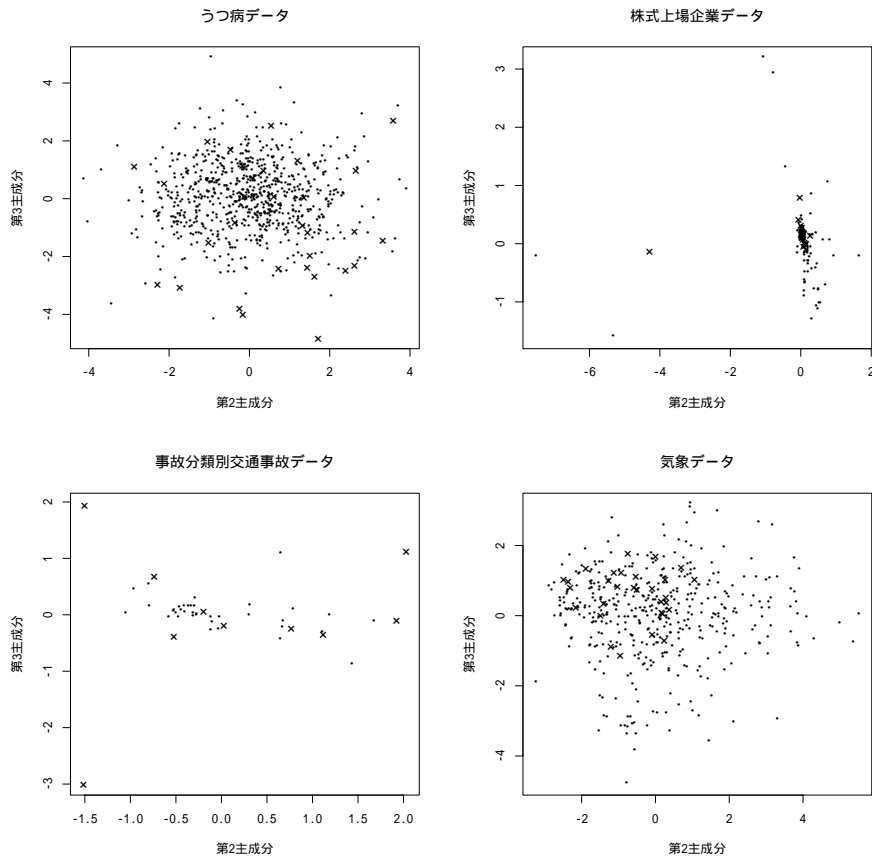


図 2 第 2 第 3 主成分プロット

主成分の和をとり，さらに，平均 0，標準偏差 1/10 の正規乱数を誤差として各変数に加える。

2.  $n$  行 8 列の負例事象群データを発生させる。発生手順は正例事象群データの場合とほぼ同様だが，第 1 主成分（または第 3 主成分）の第 1 列から第 6 列までには 3 通りのずれを加える。
3. 発生させたデータを対応分析法の距離①にそのままかけると，SN 比の真数部分が負の数になる可能性があるので，両データにそれぞれ 5 を加算する。
4. 発生させた正例事象群データと負例事象群データを 6 個のプログラム (MTS 法の距離，主座標分析法の距離①，主座標分析法の距離②，対応分析法の距離①，対応分析法の距離②，対応分析法の距離③) にかけて，各距離での誤判別率の平均を算出する。
5. 1~4 を  $N$  回試行する。



表 7 主成分得点の符号

|   | 第 1 主成分 | 第 2 主成分 | 第 3 主成分 |
|---|---------|---------|---------|
| 1 | +       | +       | +       |
| 2 | +       | -       | +       |
| 3 | +       | +       | -       |
| 4 | +       | -       | -       |
| 5 | +       | +       | +       |
| 6 | +       | -       | +       |
| 7 | +       | +       | -       |
| 8 | +       | -       | -       |

## 7.2 シミュレーション結果と考察

第 1 主成分において、ずれの大きさが 0.5 の場合の 10 回試行シミュレーション結果を表 8 に、ずれの大きさが 0.25 の場合の 10 回試行シミュレーション結果を表 9 に、ずれの大きさが 0.0625 の場合の 10 回試行シミュレーション結果を表 10 に示す。

また、第 3 主成分において、ずれの大きさが 0.5 の場合の 10 回試行シミュレーション結果を表 11 に、ずれの大きさが 0.25 の場合の 10 回試行シミュレーション結果を表 12 に、ずれの大きさが 0.0625 の場合の 10 回試行シミュレーション結果を表 13 に示す。

(表内の\*は誤判別率の平均が最小のものを示す。)

表 8 ずれの大きさが 0.5 の場合の 10 回施行シミュレーション結果 (第 1 主成分)

|       | 1 回    | 2 回    | 3 回    | 4 回    | 5 回    | 6 回    | 7 回    | 8 回    | 9 回    | 10 回   |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| MTS 法 | 0.147  | 0.092* | 0.048  | 0.057* | 0.057  | 0.047* | 0.121  | 0.056  | 0.072  | 0.088  |
| 主座標①  | 0.129* | 0.103  | 0.044* | 0.058  | 0.052* | 0.090  | 0.113* | 0.051* | 0.069* | 0.084* |
| 主座標②  | 0.129* | 0.100  | 0.044* | 0.058  | 0.052* | 0.090  | 0.113* | 0.051* | 0.069* | 0.084* |
| 対応分析① | 0.327  | 0.103  | 0.297  | 0.279  | 0.204  | 0.261  | 0.330  | 0.208  | 0.297  | 0.312  |
| 対応分析② | 0.304  | 0.235  | 0.229  | 0.238  | 0.211  | 0.271  | 0.310  | 0.201  | 0.215  | 0.298  |
| 対応分析③ | 0.321  | 0.241  | 0.244  | 0.249  | 0.296  | 0.319  | 0.337  | 0.224  | 0.230  | 0.207  |

表 9 ずれの大きさが 0.25 の場合の 10 回施行シミュレーション結果 (第 1 主成分)

|       | 1 回    | 2 回    | 3 回    | 4 回    | 5 回    | 6 回    | 7 回    | 8 回    | 9 回    | 10 回   |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| MTS 法 | 0.287* | 0.246* | 0.201* | 0.212* | 0.233* | 0.207* | 0.245  | 0.209  | 0.253  | 0.235* |
| 主座標①  | 0.316  | 0.263  | 0.229  | 0.302  | 0.243  | 0.253  | 0.196* | 0.116* | 0.250* | 0.267  |
| 主座標②  | 0.316  | 0.261  | 0.229  | 0.302  | 0.243  | 0.253  | 0.196* | 0.165  | 0.250* | 0.267  |
| 対応分析① | 0.406  | 0.428  | 0.361  | 0.436  | 0.409  | 0.456  | 0.409  | 0.395  | 0.361  | 0.391  |
| 対応分析② | 0.414  | 0.385  | 0.379  | 0.437  | 0.383  | 0.470  | 0.396  | 0.433  | 0.422  | 0.403  |
| 対応分析③ | 0.409  | 0.448  | 0.389  | 0.450  | 0.440  | 0.479  | 0.405  | 0.398  | 0.379  | 0.403  |

表 10 ずれの大きさが 0.0625 の場合の 10 回施行シミュレーション結果 (第 1 主成分)

|       | 1 回    | 2 回    | 3 回    | 4 回    | 5 回    | 6 回    | 7 回    | 8 回    | 9 回    | 10 回   |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| MTS 法 | 0.469  | 0.451* | 0.439  | 0.468  | 0.415* | 0.474  | 0.463  | 0.481  | 0.461* | 0.455  |
| 主座標①  | 0.481  | 0.469  | 0.425* | 0.466  | 0.469  | 0.463* | 0.430* | 0.483  | 0.489  | 0.425* |
| 主座標②  | 0.465  | 0.469  | 0.425* | 0.466  | 0.469  | 0.463* | 0.430* | 0.482  | 0.489  | 0.425* |
| 対応分析① | 0.435  | 0.465  | 0.454  | 0.450  | 0.452  | 0.484  | 0.456  | 0.427* | 0.468  | 0.449  |
| 対応分析② | 0.433  | 0.478  | 0.446  | 0.438* | 0.456  | 0.497  | 0.453  | 0.465  | 0.495  | 0.431  |
| 対応分析③ | 0.420* | 0.463  | 0.454  | 0.465  | 0.468  | 0.494  | 0.480  | 0.437  | 0.489  | 0.453  |

表 11 ずれの大きさが 0.5 の場合の 10 回施行シミュレーション結果 (第 3 主成分)

|       | 1 回    | 2 回    | 3 回    | 4 回    | 5 回    | 6 回    | 7 回    | 8 回    | 9 回    | 10 回   |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| MTS 法 | 0.133* | 0.071* | 0.110* | 0.110  | 0.115* | 0.125* | 0.174  | 0.142  | 0.113  | 0.129  |
| 主座標①  | 0.154  | 0.072  | 0.121  | 0.109* | 0.133  | 0.161  | 0.171* | 0.136* | 0.112* | 0.116* |
| 主座標②  | 0.154  | 0.072  | 0.121  | 0.109* | 0.133  | 0.160  | 0.171* | 0.136* | 0.112* | 0.116* |
| 対応分析① | 0.219  | 0.229  | 0.263  | 0.223  | 0.270  | 0.294  | 0.268  | 0.223  | 0.234  | 0.255  |
| 対応分析② | 0.195  | 0.206  | 0.252  | 0.203  | 0.277  | 0.300  | 0.254  | 0.303  | 0.237  | 0.242  |
| 対応分析③ | 0.210  | 0.239  | 0.264  | 0.240  | 0.240  | 0.336  | 0.281  | 0.258  | 0.248  | 0.252  |

表 12 ずれの大きさが 0.25 の場合の 10 回施行シミュレーション結果 (第 3 主成分)

|       | 1 回    | 2 回    | 3 回    | 4 回    | 5 回    | 6 回    | 7 回    | 8 回    | 9 回    | 10 回   |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| MTS 法 | 0.285  | 0.238* | 0.293  | 0.281  | 0.287* | 0.313* | 0.319* | 0.277* | 0.272* | 0.298  |
| 主座標①  | 0.271  | 0.291  | 0.292  | 0.259* | 0.322  | 0.341  | 0.337  | 0.299  | 0.276  | 0.280* |
| 主座標②  | 0.270* | 0.291  | 0.291* | 0.259* | 0.322  | 0.341  | 0.337  | 0.299  | 0.276  | 0.280* |
| 対応分析① | 0.402  | 0.384  | 0.430  | 0.374  | 0.427  | 0.436  | 0.403  | 0.356  | 0.418  | 0.413  |
| 対応分析② | 0.393  | 0.399  | 0.412  | 0.349  | 0.424  | 0.496  | 0.402  | 0.366  | 0.423  | 0.420  |
| 対応分析③ | 0.394  | 0.403  | 0.443  | 0.350  | 0.439  | 0.492  | 0.396  | 0.345  | 0.438  | 0.378  |

表 13 ずれの大きさが 0.0625 の場合の 10 回施行シミュレーション結果 (第 3 主成分)

|       | 1 回    | 2 回    | 3 回    | 4 回    | 5 回    | 6 回    | 7 回    | 8 回    | 9 回    | 10 回   |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| MTS 法 | 0.475  | 0.443  | 0.406* | 0.451  | 0.442* | 0.474* | 0.440* | 0.444* | 0.466* | 0.437  |
| 主座標①  | 0.478  | 0.441  | 0.435  | 0.456  | 0.443  | 0.476  | 0.468  | 0.477  | 0.476  | 0.417  |
| 主座標②  | 0.478  | 0.441  | 0.435  | 0.456  | 0.443  | 0.476  | 0.468  | 0.477  | 0.476  | 0.417  |
| 対応分析① | 0.439  | 0.456  | 0.449  | 0.443* | 0.462  | 0.499  | 0.447  | 0.456  | 0.467  | 0.446  |
| 対応分析② | 0.439  | 0.438* | 0.447  | 0.473  | 0.452  | 0.497  | 0.440* | 0.456  | 0.480  | 0.461  |
| 対応分析③ | 0.435* | 0.472  | 0.467  | 0.472  | 0.477  | 0.494  | 0.476  | 0.452  | 0.484  | 0.415* |

以上のシミュレーション結果をどの方法が最も優れていたかでまとめると表 14 となる。なお、各主成分のずれにおける最小値の合計回数が 10 回にならないのは、誤判別率の平均がいくつかの距離で等しい値を示したからである。

表 14 シミュレーション結果のまとめ

|        | 第 1 主成分 |      |        | 第 3 主成分 |      |        |
|--------|---------|------|--------|---------|------|--------|
|        | 0.5     | 0.25 | 0.0625 | 0.5     | 0.25 | 0.0625 |
| MTS 法  | 3       | 7    | 3      | 5       | 6    | 6      |
| 主座標分析① | 7       | 3    | 4      | 5       | 2    | 0      |
| 主座標分析② | 7       | 2    | 4      | 5       | 4    | 0      |
| 対応分析①  | 0       | 0    | 1      | 0       | 0    | 1      |
| 対応分析②  | 0       | 0    | 1      | 0       | 0    | 2      |
| 対応分析③  | 0       | 0    | 1      | 0       | 0    | 2      |

その結果、次のような傾向があることが言える。

- MTS 法の距離は、主成分のずれが大きいとき有効性が高く、ずれが小さいとき有効性がやや高い。
- 主座標分析の距離は、第 1 主成分が大きい場合に有効性が高く中程度から小さい場合にやや低い。第 3 主成分が大きいか中ぐらゐの場合に有効性がやや高く小さい場合に低い。
- 対応分析法の距離は、主成分のずれが大きい場合は有効性がかなり低く、小さい場合は有効性がやや低い。

このことから、現存する MTS 法の距離以外の距離もデータによっては有効性があり得ることが確認できた。

## 8 まとめ

本研究では、正例と負例とを判別するのに MTS 法の距離、主座標分析法の距離①、主座標分析法の距離②、対応分析法の距離①、対応分析法の距離②、対応分析法の距離③の 6 種類の距離を用いた方法を採用した。これら 6 種類の距離を用いた方法による、4 つの実データ（うつ病データ、株式上場企業データ、事故分類別交通事故データ、気象データ）に適用した検証実験結果および主成分を用いた 6 パターンのシミュレーション結果より、次のことが言える。

MTS 法の距離は基本的に推奨されることが多いが、主成分のずれが大きい場合や小さい場合は状況に応じてその推奨性が下がることもある。主座標分析法の距離は主成分のずれが大きな場合にやや推奨され、小さい場合は状況に応じて推奨される。しかし、対応分析法の距離は主成分のずれが大きな場合も小さな場合もあまり推奨されない。すなわち、今までは現存していた MTS 法の距離による方法が一般的であったが、吉野ら (2004) で提案されている主座標分析法の距離による方法も場合によっては適用できるのではないかと思われる。

## 9 おわりに

本研究では、変数選択には直交表および要因効果図を適用し、閾値決定法には正例と負例の誤判別率の平均を適用することにより、田口玄一博士によって提案された MTS 法の全自動化を統計解析ソフト R 上で実現することができた。また、吉野ら (2004) で提案された MTS 法の距離以外の距離を用いた方法の全自動化も実現することができた。そして主成分を用いたシミュレーションによって、各距離における分析法の有効性も確認することができた。

## 参考文献

- [1] 警察庁・国土交通省 (2004) : 交通安全マップ, <http://www.kotsu-anzen.jp>.
- [2] 気象庁 (2006) : 毎日の全国データ一覧表,  
<http://www.data.kishou.go.jp/synopday/index.html>.
- [3] 中津川雅史・大内東 (2000) : マハラノビス距離の分布推定に基づく閾値決定法,  
“品質工学会第 8 回品質工学研究発表大会論文集”, 18-21.
- [4] 田口玄一 (1988) : “開発・設計段階の品質工学”, 日本規格協会 .
- [5] 田中豊・脇本和昌 (1983) : “多変量統計解析法”, 現代数学社 .
- [6] 立林和夫 (2004) : “入門タグチメソッド”, 日科技連出版社.
- [7] 東京証券取引所 (2005) : 上場会社検索,  
[http://www.tse.or.jp/cgi-bin/compsearch/j\\_search](http://www.tse.or.jp/cgi-bin/compsearch/j_search).
- [8] 吉野睦・松田眞一・仁科健 (2004) : MTS 法における距離算出方法の改善,  
“日本品質管理学会第 75 回研究発表会発表要旨集”, 61-64.