

MTS 法と各距離における分析法の比較

M2005MM033 棚橋 誠

指導教員 松田 眞一

1 はじめに

現在、新聞やテレビなどの報道により、企業における品質問題が多々発生しているのを目にする。最悪の場合、人の命に関わりかねないような品質問題をなくすることはできないのだろうか。今、品質管理の重要性が問われている。

1.1 品質管理とは

「JIS Z 8101 品質管理用語」において、品質管理は以下のように定義されている。(仁科ら [1] 参照)

品質管理 (Quality Control): 買い手の要求に合った品質の品物またはサービスを経済的に作り出すための手段の体系。

品質管理を効果的に実施するためには、市場の調査、研究・開発、製品の企画、設計、生産準備、購買・外注、製造、検査、販売及びアフターサービス並びに財務、人事、教育など企業活動の全段階にわたり経営者を始め管理者、監督者、作業員など企業の全員の参加と協力が必要である。

1.2 研究目的

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。品質問題が発生してしまつたら、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。

この MTS 法をベースとして、本研究では以下の 2 点に焦点を充てる。

1. プログラム作成により、統計解析ソフト R 上での MTS 法の全自動化を実現させる。
2. MTS 法と吉野ら [5] で提案されている主座標分析法や対応分析法を含む MTS 法以外の距離による方法との比較を行うことにより、各方法の長所や短所を考察し、正常な個体と異常な個体を分類する方法の最良化を図る。

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を中心を求めて、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。(田口 [2], 立林 [4] 参照)

2.1 MTS 法の距離

表 1 正例事象群データ

事象番号	項目			
	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	\cdots	\vdots
n	x_{n1}	x_{n2}	\cdots	x_{nk}

表 2 負例事象群データ

事象番号	項目			
	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	\cdots	\vdots
m	y_{m1}	y_{m2}	\cdots	y_{mk}

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。

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

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

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。

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

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

規準化されたデータを表 3, 表 4 に示す。

表 3 正例事象群を規準化したデータ

事象番号	変数			
	u_1	u_2	\dots	u_k
1	u_{11}	u_{12}	\dots	u_{1k}
2	u_{21}	u_{22}	\dots	u_{2k}
\vdots	\vdots	\vdots	\vdots	\vdots
n	u_{n1}	u_{n2}	\dots	u_{nk}

表 4 負例事象群を規準化したデータ

事象番号	変数			
	v_1	v_2	\dots	v_k
1	v_{11}	v_{12}	\dots	v_{1k}
2	v_{21}	v_{22}	\dots	v_{2k}
\vdots	\vdots	\vdots	\vdots	\vdots
m	v_{m1}	v_{m2}	\dots	v_{mk}

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。

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

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。

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

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

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 μ を求める。

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

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 法の距離及び 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。(田中ら [3] 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。

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

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

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、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 [5] 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」と相関係数行列を用いる方法「主座標分析法の距離」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 $x_{i.}$ 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 $D_{x_i}^2$ は以下のような方法で求められる。

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

負例も正例を基準に以下のように求める。

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

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 [5] 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」と偏差値変換を用いる方法「対応分析法の距離」と重心化を用いる方法「対応分析法の距離」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 水準を判定に使用する変数、第 2 水準を判定に使用しない変数とする。
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。

R 上での各プログラム実行方法は、正例事象群データと負例事象群データを用いて以下のように入力する。

関数名 (正例事象群データ, 負例事象群データ)

各距離における関数は表 5 の通りである。

表 5 各距離における関数

距離	関数
MTS 法の距離	mts
主座標分析法の距離	pcs1
主座標分析法の距離	pcs2
対応分析法の距離	cas1
対応分析法の距離	cas2
対応分析法の距離	cas3

5 検証実験

5.1 使用データ

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データ、東京証券取引所ホームページから検索した建設業における上場企業の 2005 年度決算データ、交通安全マップから作成された事故分類別交通事故データ、気象庁ホームページから検索した 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データの 4 つのデータを使用する。

5.2 実験結果と考察

うつ病データと株式上場企業データおよび事故分類別交通事故データではどれも MTS 法が最良となり、誤判別率の平均はそれぞれ 0.118, 0.224, 0.141 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ので、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てからの実験も行なった。その結果、主座標分析法の距離は事故分類別交通事故データに対し MTS 法の距離と同等の機能を示した。一方、気象データでは対応分析法の距離が最良となり、誤判別率の平均は 0.466 であった。

このことから、データによっては従来の MTS 法の距離を適用するよりも吉野ら [5] によって提案されている MTS 法の距離以外の距離 (主座標分析法の距離, 対応分析法の距離) を適用した方が、正例と負例を分類するのに適していることもあるということが考察できる。

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 主成分プロットをそれぞれ図 1 に示す。

ただし、各プロット図において、 \cdot は正例データ、 \times は負例データを示す。

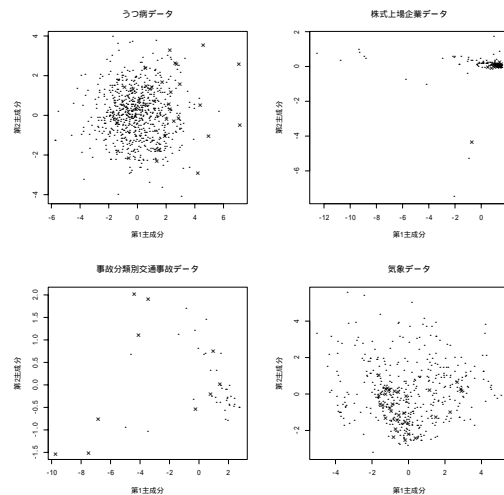


図 1 第 1 第 2 主成分プロット

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上場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通りの大きさに発生させて、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、各主成分が直交する条件で 3 つの主成分の和をとり、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 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ため、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、主座標分析法の距離、対応分析法の距離、対応分析法の距離、対応分析法の距離) につけ、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。

7.2 シミュレーション結果と考察

10 回施行のシミュレーションを実行した。各主成分のずれにおける誤判別率の平均が最小値を示した回数は表 6 の通りである。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 6 シミュレーション結果

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

このことから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い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、主座標分析法の距離、対応分析法の距離、対応分析法の距離、対応分析法の距離の 6 種類の距離を適用した方法を採用した。これら 6 種類の距離を適用した方法による、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) を適用した検証実験結果および主成分を適用した 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 [5] 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 [5] 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。そして主成分分析法を適用したシミュレーションによって、データの種類による各距離を適用した方法の有効性も確認することができた。

参考文献

- [1] 仁科健, 澤田善次郎, 伊藤賢次: 品質管理, 日刊工業新聞社 (1995).
- [2] 田口玄一: 開発・設計段階の品質工学, (財) 日本規格協会 (1988).
- [3] 田中豊, 脇本和昌: 多変量統計解析法, 現代数学社 (1983).
- [4] 立林和夫: 入門タグチメソッド, 日科技連出版社 (2004).
- [5] 吉野睦, 松田眞一, 仁科健: MTS 法における距離算出方法の改善, (社) 日本品質管理学会第 75 回研究発表会発表要旨集, pp.61-64 (2004.9).