

MTシステムの多重共線性の対策に関する研究

KTS&C
菊地 富男

1. はじめに

MTシステムで判別や診断の用途に使用されている方法は、MT法(逆行列)、MTA法(余因子)、MTS法(シュミット)である。MTシステムには、「多重共線性(通称、マルチコ)」と呼ばれている問題があり、広く使われているMT法では、逆行列が計算できないという現象が発生する。

そこで、逆行列の演算に新しい方式(一般化逆行列)を導入することにより、MT法の問題を解消することができるかどうかについて調査した以下の結果について報告する。

- (1) 多重共線性の一般的な定義について
- (2) 逆行列を求める方法について
- (3) 一般化逆行列法と他の方法との優劣比較について

2. 多重共線性の一般的な定義について

「多重共線性」という言葉は、多変量解析の中の重回帰分析や主成分分析などに関連し、統計学で使用されている用語である。重回帰分析において、独立変数間に強い相関がある場合、または強い一次従属関係(ある変数が別の変数の線形結合になっていること:例えば、「 $x_2 = 2x_1 + 3$ 」)になっていること)がある場合を「多重共線性(multi-collinearity)の問題がある」という。

多重共線性には、「正確多重共線性」と「準多重共線性」がある(詳細は、前報「MTシステムにおける多重共線性に関する研究」を参照)。

一次従属の列(行)があると、階数(ランク) = $n < p$ (変数の数) となり、行列式「 $|R|$ 」(determinant:あるいはDET)は、「0」となる。階数(ランク) = p の行列を正則(regular)といい、逆行列が存在する。正則でない行列は、特異(singular)といい逆行列は存在しない。 $n < p$ の多変量データ行列は、特異になる。相関係数行列の行列式 $|R|$ は、範囲(0, 1)となり、正確な一次従属関係の時「0」、相互に直交していれば「1」になる。「0」に接近するにつれて多重共線性の程度が大きくなるので、 $|R|$ は多重共線性の指標として利用できる。なお、階数(ランク)とは、一次独立な列(行)の最大数である。

$$\text{相関係数行列 } (R) = \begin{vmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2p} \\ & & \cdots & & \\ r_{p1} & r_{p2} & r_{p3} & \cdots & r_{pp} \end{vmatrix} \quad (1)$$

$$\text{余因子行列 } (\tilde{R}) = \begin{vmatrix} A_{11} & A_{21} & A_{31} & \cdots & A_{p1} \\ A_{12} & A_{22} & A_{32} & \cdots & A_{p2} \\ & & \cdots & & \\ A_{1p} & A_{2p} & A_{3p} & \cdots & A_{pp} \end{vmatrix} \quad (2)$$

$$\text{逆行列 } (R^{-1}) = \frac{\tilde{R}}{|R|} \quad (3)$$

$r_{11} \sim r_{pp}$: 項目間の相関係数 , R : 相関係数行列 , R^{-1} : 相関係数行列の逆行列

\tilde{R} : 相関係数行列の余因子行列 , $|R|$: 行列式(DET)

余因子行列(式(2))(A_{11} (1行、1列)の余因子行列)とは、相関係数行列「 R 」の1行および1列の要素を除いた余った要素で作る行列である。

(1) 正確多重共線性

変数(項目)の数「p」の相関係数行列R(式(1))において、階数(ランク)が「p」ではなく、「p-1」、「p-2」、…、「p-r」のとき(r個のランク落ちがあるという)または、サンプル数「n」と項目数「p」が「n ≤ p」であれば正確多重共線性が起きる。このとき、相関係数行列の逆行列 R^{-1} (式(3))は定義(計算)できない。

(2) 準多重共線性

正確な一次従属関係にはないが、近似的に一次従属関係がある場合、準多重共線性があるという。例えば、二つの項目間の相関係数が「1」に近い場合に発生し、行列式 $|R|$ が、「0」に限りなく近くなり、逆行列が計算できなくなったり、計算できていても誤差が大きく結果が信頼できないことが発生する。通常、「多重共線性」というと、それは、「準多重共線性」を意味する。

2. 逆行列を求める方法について

逆行列を求める方法としては、

- ・余因子行列(Ajout Matrix)を使う方法
- ・ガウス・ジョルダン法(Gauss-Jordan)を使う方法
- ・一般化逆行列(Generalized Inverse Matrix)を使う方法

などがある。最も一般的な方法は、「ガウス・ジョルダン法」であり、それを更に改良した「LU分解法」、「ドゥリットル法」、「改定コレスキ法」と呼ばれている方法もある。

MT法(逆行列)のオリジナルソフトの逆行列計算も、「ガウス・ジョルダン法」であるが、正確多重共線性あるいは準多重共線性に対して、逆行列が計算できない問題が発生する。そこで、「一般化逆行列法」を導入することにより、この問題を解消できないかどうかを検討することにした。

一般化行列法は、数学あるいは数値計算法の中では、広く使われている方法であり、行列が正方行列でない場合(例えば、 $n \neq m$ の n 行 \times m 列の行列)や正則でない特異行列の近似的な逆行列(R^+)を求める方法である。

一般化行列法の中で一意性のある(解が一意に決まる)「ムーア・ペンローズ(Moore-Penrose)逆行列の特異値分解」と呼ばれている方法を採用する。

(1) 主成分分析における固有値と固有ベクトルについて

一般化逆行列法では、「固有値」と「固有ベクトル」を用いるが、多変量解析に分類されている「主成分分析」と関係が深い統計量であり、ここで、「主成分分析」について、簡単に説明する。

「主成分分析」とは、数多くある変数(正確には、説明変数)の中から、「○○○総合力」「×××能力」といった新しい概念のファクターを導く方法である。

このファクターをスケールとして、変数や個々のサンプルの類似性あるいはポジショニングを明らかにすることができる。主成分分析を行うと一般的には、結果は次の様に表現される。

$$\begin{cases} Z_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \\ Z_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \\ \vdots \\ Z_m = a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mp}x_p \end{cases} \quad (4)$$

この(4)式を主成分といい、 Z_1 を第一成分、 Z_2 を第二主成分…といい、式の係数、「 a_{11} 、 a_{12} 、…、 a_{1p} 、…、 a_{mp} 」を「固有ベクトル」という。言い換えれば、主成分分析は、 p 個の変数 x_1, x_2, \dots, x_p を m 個の新しいファクター、すなわち、「主成分 Z_1, Z_2, \dots, Z_m 」に集約する方法である。したがって、「 p 」と「 m 」とは、「 $p \geq m$ 」の関係にある。

また、次に述べる「固有値($\lambda_1, \lambda_2, \dots, \lambda_m$)」は、新しいファクター「主成分 Z_1, Z_2, \dots, Z_m 」の重要度を表し、重要なファクターほど固有値が大きな値となる。

(2) ムーア・ペンローズの一般化逆行列 R^+ について

相関行列が正則な(逆行列が計算できる)とき、そのスペクトル分解を

$$R = \lambda_1 \omega_1 \omega_1^T + \lambda_2 \omega_2 \omega_2^T + \dots + \lambda_p \omega_p \omega_p^T \quad (5)$$

λ_i : 固有値 , ω_i : 固有ベクトル ()^T : ()行列の転置行列(行と列を入れ替えた行列)と表す。その逆行列 R^{-1} のスペクトル分解は,

$$R^{-1} = \frac{1}{\lambda_1} \omega_1 \omega_1^T + \frac{1}{\lambda_2} \omega_2 \omega_2^T + \dots + \frac{1}{\lambda_p} \omega_p \omega_p^T \quad (6)$$

ここで、相関係数行列がランク落ちしている場合(逆行列が計算できない場合)を考える。そのランクを「 $q(q < p)$ 」とする。このときのスペクトル分解は一般性を失うことなく,

$$R = \lambda_1 \omega_1 \omega_1^T + \lambda_2 \omega_2 \omega_2^T + \dots + \lambda_q \omega_q \omega_q^T \quad (7)$$

と表現できる。上式(7)に現れる固有値は全て正值としている。このとき、逆行列に代わるものとして一般化逆行列を導入する。通常的一般化逆行列は一意的でないので、一意性のあるムーア・ペンローズの一般化逆行列「 R^+ 」を用いると、それは,

$$R^+ = \frac{1}{\lambda_1} \omega_1 \omega_1^T + \frac{1}{\lambda_2} \omega_2 \omega_2^T + \dots + \frac{1}{\lambda_q} \omega_q \omega_q^T \quad (8)$$

である。ところで、ランク落ちした状況では、(6)式の右辺には、「 $0/0$ 」の不定形が、「 $p-q$ 」個現れる。そこで、固有値が正の値の項だけを取り出すと、それは逆行列「 R^{-1} 」の代わりにムーア・ペンローズの一般化逆行列「 R^+ 」を用いる(8)式となる。相関係数行列がランク落ちしなければ、「 R^{-1} 」と「 R^+ 」とは等しい。

3. 一般化行列法の他の方法との優劣比較について

従来方式(ガウス・ジョルダン法:以下「ガウス法」と称す)、新しい方式(一般化逆行列法:以下「一般化法」と称す)およびMTS法について、正確多重共線性がある場合および準多重共線性がある場合について、優劣を比較した結果を以下に示す。

3.1 正確多重共線性

(1)階級(ランク)落ちが発生した場合

項目(変数)が10項目($X1 \sim X10$)の多重共線性がない事例を使って、正確多重共線性をもつ、第11項目($X11$)「 $X11 = X9 + X10$ 」を追加し、問題のない10項目のみで解析した結果と第11項目を追加した正確多重共線性が発生している場合の結果を表1、表2(表の1~10は、単位空間を構成するサンプル10個のマハラノビス距離 MD^2)に示す。

- 1) ガウス法は、逆行列が計算できないことが発生する。
- 2) 一般化法は、第11項目を追加する前の10項目のみの結果と同じになる(正確多重共線性があると階級(ランク)落ちが発生し、第11項目を削除した結果と同じになる)。
- 3) MTS法は、計算もでき、かつその結果も正常時と近似している。

表1 正常場合の単位空間のマハラノビス距離

	1	2	3	4	5	6	7	8	9	10
ガウス法	0.876	0.920	1.436	1.108	0.416	1.340	0.690	0.580	1.226	0.633
一般化法	0.876	0.920	1.436	1.108	0.416	1.340	0.690	0.580	1.226	0.633
MTS法	0.876	0.920	1.436	1.108	0.416	1.340	0.690	0.580	1.226	0.633

表2 正確多重共線性(ランク落ち)の単位空間のマハラノビス距離

	1	2	3	4	5	6	7	8	9	10
ガウス法	逆行列が計算できない									
一般化法	0.876	0.920	1.436	1.108	0.416	1.340	0.690	0.580	1.226	0.633
MTS法	0.907	0.999	1.387	1.013	0.389	1.254	0.629	0.547	1.141	0.678

(2)項目数(p) \geq サンプル数(n)の場合

項目(変数)が10項目($X1 \sim X10$)で、問題のない単位空間のサンプル数を12個とした場合および10個とした場合の結果を表3、表4(表の1~10は、単位空間を構成するサンプル10個のマハラノビス距離 MD^2)に示す。

- 1) ガウス法は、逆行列が計算できないことが発生する。
- 2) 一般化法は、計算はできるが、単位空間および異常品のマハラノビス距離(MD^2)が、すべて「1」となり、

単位空間と異常品とのパターン差を検出できない。

3) MTS法は、計算もでき、かつその結果も正常時と近似している。

表3 正常場合の単位空間のマハラノビス距離

	1	2	3	4	5	6	7	8	9	10
ガウス法	0.912	1.096	1.072	1.061	0.817	1.050	1.036	0.724	1.098	0.956
一般化法	0.912	1.096	1.072	1.061	0.817	1.050	1.036	0.724	1.098	0.956
MTS法	0.912	1.096	1.072	1.061	0.817	1.050	1.036	0.724	1.098	0.956

表4 正確多重共線性(p≥n)の単位空間のマハラノビス距離

	1	2	3	4	5	6	7	8	9	10
ガウス法	逆行列が計算できない									
一般化法	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MTS法	0.909	0.958	0.900	1.056	1.245	0.916	0.911	1.256	0.930	0.919

3. 2 準多重共線性

1項目目は実際のデータを使用し、2項目目を1項目目のデータにエクセルの一樣乱数を $1/10^6 \sim 1/10^2$ 倍して加え、項目間の相関係数の値を加減しながら、相関係数(r)、行列式(|R|)および単位空間のサンプル10個のマハラノビス距離(MD²)を評価する。その結果を表5(表の1~10は、単位空間を構成するサンプル10個のマハラノビス距離MD²)に示す。

(1)ガウス法は、相関係数が0.9999999以上になると逆行列が計算できない。

(2)一般化法は、相関係数が0.999999以下の場合、他の方法と同じ結果となるが、相関係数が0.999999以上になると階級(ランク)落ちが発生し、多重共線性に関係している項目を削除した結果と同じとなる。

(3)MTS法は、計算上の問題もなく、その結果も信頼できる(今回使用している解析法以外の方法で解析した結果と一致する)。

表5 準多重共線性の評価結果

相関係数	r	0.51119539870
行列式	DET	0.738679264

	1	2	3	4	5	6	7	8	9	10
ガウス法	1.049	0.906	2.574	1.075	2.663	1.303	1.377	0.330	0.833	1.749
一般化法	1.049	0.906	2.574	1.075	2.663	1.303	1.377	0.330	0.833	1.749
MTS法	1.049	0.906	2.574	1.075	2.663	1.303	1.377	0.330	0.833	1.749

相関係数	r	0.98860527387
行列式	DET	0.022659612

	1	2	3	4	5	6	7	8	9	10
ガウス法	1.071	1.585	2.198	0.321	3.070	1.181	0.559	0.603	0.012	1.592
一般化法	1.071	1.585	2.198	0.321	3.070	1.181	0.559	0.603	0.012	1.592
MTS法	1.071	1.585	2.198	0.321	3.070	1.181	0.559	0.603	0.012	1.592

相関係数	r	0.99988780369
行列式	DET	0.00022438

	1	2	3	4	5	6	7	8	9	10
ガウス法	0.677	2.398	2.616	0.204	2.013	2.190	0.162	0.195	0.002	1.615
一般化法	0.677	2.398	2.616	0.204	2.013	2.190	0.162	0.195	0.002	1.615
MTS法	0.677	2.398	2.616	0.204	2.013	2.190	0.162	0.195	0.002	1.615

相関係数	r	0.99999878423
行列式	DET	2.43153E-06

	1	2	3	4	5	6	7	8	9	10
ガウス法	0.420	1.757	2.544	0.676	2.129	1.661	0.159	0.725	0.725	2.820
一般化法	0.420	1.757	2.544	0.676	2.129	1.661	0.159	0.725	0.725	2.820
MTS法	0.420	1.757	2.544	0.676	2.129	1.661	0.159	0.725	0.725	2.820

相関係数	r	0.99999998950
行列式	DET	2.10062E-08

	1	2	3	4	5	6	7	8	9	10
ガウス法	0.277	1.014	2.221	0.197	2.507	0.696	0.075	0.478	0.408	1.782
一般化法	0.552	1.787	3.642	0.255	3.643	1.159	0.088	0.286	0.005	3.178
MTS法	0.277	1.014	2.221	0.197	2.507	0.696	0.075	0.478	0.408	1.782

相関係数	r	0.9999999989
行列式	DET	2.1108E-10

	1	2	3	4	5	6	7	8	9	10
ガウス法					計算できない					
一般化法	0.552	1.787	3.643	0.255	3.643	1.159	0.088	0.286	0.004	3.178
MTS法	1.277	1.394	2.590	0.427	1.821	1.573	0.726	0.567	0.082	1.666

4. 結果に対する考察

(1) 正確多重共線性について

・変数(項目)の数「p」とサンプル数「n」が、 $p < n$ の関係にあり、項目間に一次従属関係が成り立っている場合は、その制約式の数分だけ、階級(ランク)落ちが発生する。従来の「ガウス・ジョルダン法」では、ランク落ちが発生した場合、理論的にも逆行列は定義(計算)できないが、新しい「一般化逆行列法」では、「固有値」「固有ベクトル」を求め、有効な「固有値」のみを使って、擬逆行列(近似的な逆行列)を求めているため、計算できない問題は発生しない。

「一般化逆行列法」は、マニュアルで正確多重共線性に関係している項目を調査し、一部の項目を削除する工数を削減でき、その結果も信頼できる。

・変数(項目)の数「p」とサンプル数「n」が、 $p \geq n$ の関係にあるいわゆる特異行列の場合は、新しい「一般化逆行列法」では、計算はできるが、単位空間および異常品のマハラノビスの距離(MD^2)が、全て「1」になり、単位空間と異常品のパターン差を検出できない。

(2) 準多重共線性について

従来の「ガウス・ジョルダン法」では、2項目間の相関係数が「0.999999」以上の場合は、正確多重共線性と同じ様に、逆行列が計算できないことが発生する。

新しい「一般化逆行列法」では、2項目間の相関係数が「0.999999」以上になると正確多重共線性と同じ様に階級(ランク)落ちが発生し、「ガウス・ジョルダン法」や「MTS法」の解析結果と異なっている。

これは、2項目間の相関係数が「0.999999」以上になると二つの変数は同じものであると判断し、片方の項目を削除して解析するのに等しい。

「一般化逆行列法」では、「固有値」をどこまで有効として、擬逆行列を求めるかをソフト上で定数として設定する必要があり、今回使用しているソフトは、その設定を「 10^{-12} 」としている。この設定により、正確多重共線性および準多重共線性が発生したとき、階級(ランク)落ちが発生する値が変動する。

(3) ガウス・ジョルダン法と一般化逆行列法の特徴について

ガウス・ジョルダン法と一般化逆行列法の特徴をまとめた表を表6に示す。

表6 ガウス・ジョルダン法と一般化逆行列法の特徴

項目	ガウス・ジョルダン法	一般化逆行列法
一次従属	×	○
$p \geq n$	×	△
準多重共線性	×	○

○：一般化行列が計算でき、その結果も信頼できる

△：一般化行列が計算できるが、その結果が信頼できない

×：逆行列が計算できないあるいは計算できないことが発生する可能性がある

5. まとめ

(1)「一般化逆行列法」は、正確多重共線性および準多重共線性に対する対策として有効であり、その解析結果も信頼できる。

(2)MT法(逆行列)のオリジナルソフトに「一般化逆行列法」を追加し、従来法と併用することによって、更なるノウハウの習得とブラッシュアップを図ってゆく。

参考文献

(1) 兼高武貳他:MTシステムにおける技術開発, 日本規格協会(2002)

(2) 田口玄一:品質工学の数理, 日本規格協会(1999)

(3) コンピュータによる情報設計の技術開発, 日本規格協会(2004)

(4) 宮川雅巳:品質を獲得する技術, 日本科学技術連盟(2000)

(5) 宮川雅巳他:マハラビス・タグチ・システムにおける多重共線性対策について, 日本品質管理協会(2003)