

中心機能バイナリーデータ行列の因子分析における問題点

——中心集落の階層構造分析への適用において——

碓 井 照 子*

On the Problem for the Application of a Factor Analysis to the Central Function-binary Data Matrix

Teruko Usui

(1976年9月30日受理)

はじめに

規模におけるそれぞれの段階(低次→高次)に応じて、中心集落は低次の中心機能から高次の中心機能を有している。それ故、中心機能の分類という観点から中心集落の階層構造を分析した研究例は数多くあるが¹⁾、その分類方法は研究者によって様々であり、従来、その曖昧性ゆえに批判されてきた。唯一の客観的分類法は、Berry が利用した直接因子分析法 (Direct Factor Analysis) である²⁾。中心機能の分類に因子分析を適用したことは高く評価されるが、彼が使用した分析法には問題が多い。直接因子分析法は、因子分析法の中でも特殊なもので、相関行列 (Correlation Matrix) を使用せず、直接、データ行列 (Data Matrix) に因子分析を適用するものである³⁾。現在、一般的に利用されている因子分析法は、データの相関行列から因子負荷行列 (Factor loading Matrix) を求めようとするものであり、相関分析法の一種であるとも言われている⁴⁾。Berry はデータ行列として Incidence Matrix⁵⁾ を使用しているが、これは、各々の中心機能が各々の中心集落に立地しているか否かを1.0, 0.0のバイナリー型で表現した行列である。もともと、因子分析法は量的変数のために発展されてきたものであり、バイナリー型の質的変数に因子分析を適用することが、特殊なのである。更に、Berry が利用した直接因子分析法は、Saunders の博士論文によって発表されたものであり⁶⁾、この方法はいまだ因子分析法としての評価は定っておらず、博士論文であるゆえに、論文の入手が困難である。また、Berry は、直接因子分析法のコンピュータープログラムを彼の論文⁷⁾で明示していない。それ故、この方法を利用することは、現段階において不可能に近い。筆者は、Berry が使用した Incidence Matrix と同一性質を有する中心機能バイナリーデータ行列に、Berry とは異なる一般的な因子分析を試みてみた。

バイナリーデータ行列の因子分析と言っても因子分析法それ自体は、相関行列から因子負荷行列を求める一般的な方法を利用するわけであるが、問題は、バイナリーデータ行列から相関行列を求めるところにある。バイナリーデータ行列の相関を求めるため、本研究では、最少二乗シンプレックスデータマトリックス解法 (The Least Square Simplex Data Matrix Solution)⁸⁾ を利用した (以後、この方法を、LSSDM 法と略す)。これは、バイナリーデータ行列の相関行列を求める方法としては一般的なものであるが、既に指摘

* 地理学研究室

したように、もともとバイナリーデータ行列の因子分析自体が特殊なのである。それ故、プログラムパッケージとしてこの解法をプログラム化しているコンピューターセンターは無く、Horst の示したプログラム⁹⁾を参考に、この解法の分析から始めざるを得なかった。また、Horst が、どのようなバイナリーデータ行列にこの解法が適しているかという点について、明確に説明していないゆえに、総ての作業は試行錯誤的に行わざるを得ず、究極的に、この解法は、中心機能バイナリーデータ行列の分析には不適であるということが明らかになったのである。

本論文は、何故、中心機能バイナリーデータ行列の分析において不適であったのかを、この解法の原理、計算方法を考察する中で明確にし、最後に、中心機能バイナリーデータ行列への具体的な適用例を提示したい。

1. 残差相関行列を利用した因子分析法について

(1) 方法の原理とシンプレックスバイナリー行列

周知のようにバイナリーデータ行列の各成分は、1.0か0.0のどちらかである。この行列における各変数は、量的な変数とは異なって質的な変数であり、各変数は一つ或は幾つかの項目集合やカテゴリーで表わされている。つまり、ある変数が幾つかの項目集合で表わされている場合、各項目の内容に二者択一的な選択がなされ、それぞれ1.0か0.0で表現されるのである。それ故、各変数間の相関係数を求める場合、各変数における項目の性質が相関係数に大きな影響を与える。このことは、バイナリーデータを扱う場合の問題点として早くから Ferguson などによって指摘されていたことである¹⁰⁾。例えば、ある項目が他の項目に比べて、項目それ自体に選択され易い性質を有している場合、或はまた、ある項目に対する正解を1.0それ以外を0.0で表現した場合における各項目間における難易度の問題など、項目それ自体に含まれる選好性や難易度などによって項目間の均一性にばらつきが出来てしまい、このことによって各変数間の相関係数に影響を与えるのである。各変数が全く均一な項目から成立っておれば、このような影響は生じないわけだが、現実の問題として均一な項目ということは不可能である。項目の選好性や難易度 (item preference or difficulty) などのばらつきに起因する影響を出来る限り小さくする必要があり、バイナリーデータ行列から各変数間の相関係数を求める場合、このことが最大の問題となってくるのである¹¹⁾。

この問題をどのように処理するかについて、従来から様々な方法が考え出されてきたが、これら項目の選好性や難易度などに起因する現象を如何に把えるかによって二つの方法が考えられる¹²⁾。一つは、これらの現象を厄介現象 (nuisance phenomenon)¹³⁾ と考える立場である。経験的なデータにおいては、どうしても測定誤差が生じるし、様々な原因によってでたらめな現象が付随するものである。それ故、このようなデータを処理する場合、これらの現象をデータから排除するために何らかの方法が利用される。このような方法の一つとして、四分相関係数 (Tetrachoric Correlation Coefficient)¹⁴⁾ が利用されるのである。つまり、この立場は、項目の選好性や難易度などに起因する現象を経験的なデータにみられる厄介現象と考え、四分相関係数によってバイナリーデータにおける各変数間の相関を求めようとするものである。もう一つの立場は、これらの現象をバイナリーデータ行列の次元性 (dimensionality) との関連で把え、バイナリーデータ行列の中からこの現象部分を分離することによってこれらの現象を除去してしまおうとするものである。つまり、項目の選好性や難易度などに起因する現象を厄介現象のような付随現象としてではなく、バイナ

リーデータ行列の中に本来、含まれた構造的なものとして把らえ、これらをバイナリーデータ行列の中から構造的に分離させることによって項目の選好性や難易度などに影響されないバイナリーデータ行列を見付け出そうとするものである。項目の選好性や難易度などに起因する現象を前者の方法のようになくしてしまうのではなく、構造的にとりあげ多変量解析的に処理しようとする後者の方法の方が現実的である。更に、四分相関係数は属性変数の正規分布という非現実的な仮定を条件とするため、現在では殆んど使用されていない。それ故、本論文では後者の方法を取ることにした。

Horstによると、この後者の方法の理論的基盤は Guttman のシンプレックス (Simplex) の考え方にある¹⁵⁾。1950年、Guttman が彼の Scaling Theory において、バイナリーデータ行列におけるこれらの問題に理論的な基礎を最初に提唱したのであり、これがシンプレックスの考え方である。この理論は、次のような仮説を前提とする。「ある項目に十の反応を示した総ての人は、その項目よりも選好し易い項目に関しては総て十の反応を示す。」¹⁶⁾ この仮説をもとに、項目の選好性や難易度などに起因する現象をシンプレックスバイナリー行列 (Simplex Binary Matrix) として表現しようとするものである。更に、この方法の特徴は、バイナリーデータ行列に含まれている項目の選好性や難易度などに起因する部分をシンプレックスバイナリー行列という形で抽出した後、バイナリーデータ行列からこのシンプレックスバイナリー行列を引き去って、残差バイナリー行列 (Residual Binary Matrix) を求めることにある。この残差バイナリー行列は、項目の選好性や難易度などに起因する現象を含んでいない行列であり、この行列の相関係数は、項目の諸性質に影響されることはない。この残差バイナリー行列の相関行列を残差相関行列 (Residual Correlation Matrix)¹⁷⁾ とすれば、この残差相関行列に因子分析を行うのである。

この方法において最大の問題は、バイナリーデータ行列から如何にしてシンプレックスバイナリー行列を求めるかということにある。本論文では、最も基本的な解法としての LSSDM 法を利用することにした¹⁸⁾。

(2) 最少二乗シンプレックスデータ行列解法における問題点

LSSDM 法とは、「シンプレックスバイナリー行列の列和(sums of columns), 或は行和(sums of rows) がバイナリーデータ行列の列和, 或は行和に等しい。」¹⁹⁾ という仮定のもとに、バイナリーデータ行列に最も近似したシンプレックスバイナリー行列を最少二乗法によって求めようとするものである。シンプレックスバイナリー行列とは、あくまで仮説的な行列でありデータ行列とは独立に存在するのではなく、バイナリーデータ行列の中に潜在化されている行列である。それ故、シンプレックスな構造を有し、バイナリーデータ行列の特性も兼備している。つまり、シンプレックスな構造とは、前述したバイナリーデータ行列における項目の選好性や難易度などに起因するものであり、データ行列の特性を兼備するということが、この解法の仮定なのである。

この解法の基本的な考え方は、バイナリーデータ行列とシンプレックスバイナリー行列とにおけるこのような関係を前者を従属変数 (y), 後者を独立変数 (x)²⁰⁾ とする回帰方程式 ($x\mathbf{B}-y=\epsilon$)²¹⁾ として求め、 ϵ を最少にすることによってシンプレックスバイナリー行列をバイナリーデータ行列に近似させるところにある²²⁾。質的な変数の多変量解析において、有力な方法の一つにダミー変数²³⁾があるが、最少二乗シンプレックスデータ行列解法の場合も質的な変数群からなるバイナリーデータ行列やシンプレックスバイナリー行列を一種のダミー変数とみなして定量的に処理しようとするものであり、この解法の基本的な考え方がダミー変数のそれに共通するものであると言える。

以上のように、シンプレックスバイナリ行列を求めるについて最も重要なことは、バイナリデータ行列との近似性ということである。何を基準に近似させるかが問題であり、この解法においては、前述した「両行列において列和或は行和が等しい。」という仮定がその基準になっているのである。バイナリデータ行列の列和或は行和が、この解法に与える影響は非常に大きい。次に、この点における問題点をこの解法の具体的な計算方法を考察することによって明確にしてみたい。

Horst の説明によると、具体的な計算過程において必要な関係式は下記の 8 つである²⁴⁾。

$$B = nM'M - (M'I)(I'M) \quad (1)$$

$$K = nM'L - (M'I)(I'L) \quad (2)$$

$$W = KT^{-1} \quad (3)$$

$$DI = \left[\frac{L'I}{n} \right] T^{-1} \quad (4)$$

$$H = \frac{(WD^{-1})W'}{n^2} \quad (5)$$

$$F = H + \frac{K_{.i}K_{.i}'}{n^2(1-p_i)} \quad (6)$$

$$nE = B - F \quad (7)$$

$$R = D_{\#}^{-1} E D_{\#}^{-1} \quad (8)$$

n : サンプル総数

m : 変数の数

M : バイナリデータ行列 ($n \times m$)

L : シンプレックスバイナリ行列 ($n \times m'$: $m' \leq m$)

B : バイナリデータ行列の n^2 倍の共分散行列 ($m \times m$)

K : バイナリデータ行列とシンプレックスバイナリ行列との n^2 倍の共分散行列 ($m \times m$)

T : 主対角線よりも下の成分が総て 1 である下三角行列 (Lower Triangular Matrix) ($m \times m$)

$P : \frac{M'I}{n} = P$ バイナリデータ行列の i 列和を n で割った p_i を成分とする列

ベクトル ($i = 1, m$)²⁵⁾

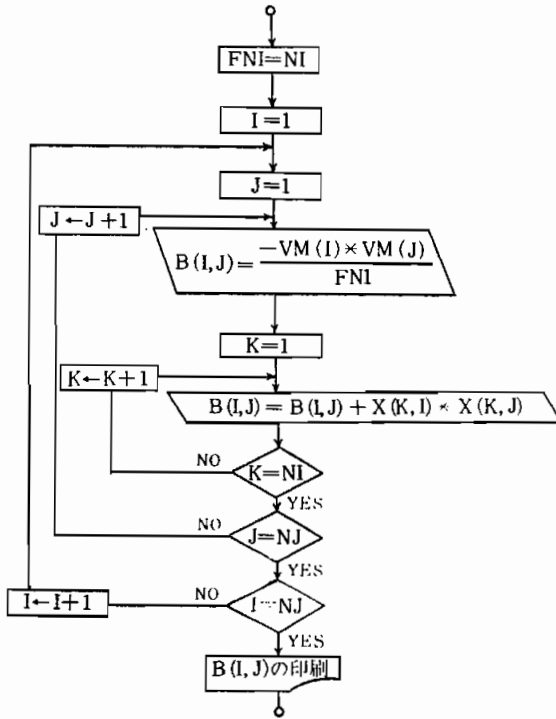
E : 残差共分散列より ($m \times m$)

R : 残差相関行列 ($m \times m$)

$D_{\#}$: 行列 E の対角成分

I : 単位ベクトル

関係式(1)から(8)までを順次計算することによって残差相関行列が求められるわけであるが、Horst のプログラムによるとコンピューターを利用しての現実の計算では、行列 B , W , H , F , E , R , 列ベクトル DI のみを求めており、行列 W や H を求めるために関係式(2)→(3)→(4)という手続きを踏まえていない。つまり、関係式(2), (3), (4)は、未知なるシンプレックスバイナリ行列 (L) をその関係式の項として含んでいるからであるが、Horst はこのことについて簡単な説明²⁶⁾を与えているだけで、特に行列 W の計算法については不明確な箇所が多い。結局、Horst のプログラムにおける計算の流れを分析することによって明確にする以外方法はなく、彼のプログラムに様々なデータを入力して行列 W の計算法を分析してみた。



NI : サンプル数
 NJ : 変数の数
 NL : シプレックスバイナリー行列の変数の数
 B(NJ,NJ) : NI倍されたバイナリーデータ行列の共分散行列
 X(NI,NJ) : バイナリーデータ行列
 VM(NJ) : データ行列の列和

図1 行列Bのフローチャート

8つの関係式において、バイナリーデータ行列 (M), サンプル総数 (n), 変数の数 (m), 下三角行列 (T) は既知のものであり、列和の平均 (P) はバイナリーデータ行列 (M) より容易に求めることができる。しかし、それ以外のものは、これらの関係式によって順次求めていかざるを得ない。まず、最初の計算は関係式(1)によるとバイナリーデータ行列の共分散行列の n^2 倍を求めるのであるが、プログラムでは、図1に示したフローチャートを見てもわかるようにバイナリーデータ行列の共分散行列の n 倍を計算している。次に行列 K, W, 列ベクトル DI の計算であるが、これらは未知であるシプレックスバイナリー行列 (L) をその項として含んでいるため、このままの形で計算することは不可能である。つまり、シプレックスバイナリー行列 (L) の諸特性からこれらの関係式を計

算可能な形に変換しなければならない。シプレックスバイナリー行列 (L) の諸特性とは、前述した(1), バイナリーデータ行列の列和, 或は行和がシプレックスバイナリー行列のそれらに等しい。(Horst の計算法では列和が等しい)。(2), Guttman が指摘したシプレックスな構造を有している。の二点である。特性(1)を数式で表現すると

$$I'M = I'L = nP' \tag{9}$$

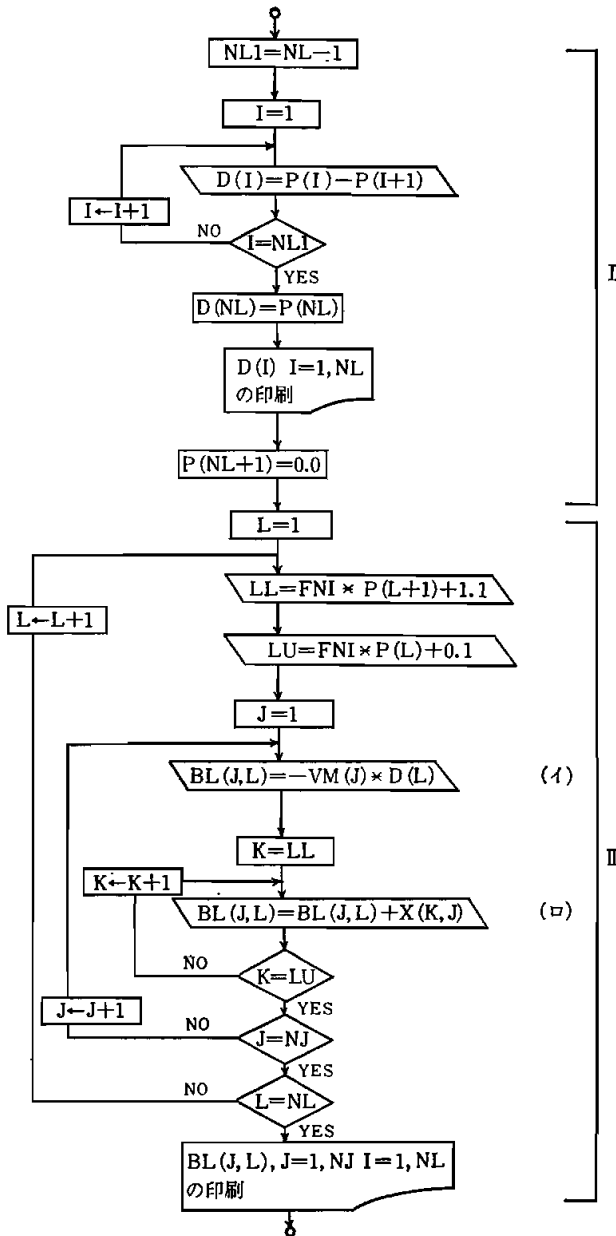
になる²⁷⁾。特性(2)は、Guttman のシプレックスの定義を数学的に表現したものである。それ故、シプレックスバイナリー行列 (L) において、この行列の積和は次のような値をとる。「シプレックスバイナリー行列において、i 列の列和の方が j 列の列和より小さいとすると、i 列と j 列の積和は i 列の列和に等しい。」この行列の積和は、常に列和の小さい方の値をとる。Horst はこの関係を次のような関係式で表現している²⁸⁾。

$$\frac{L'L}{n} = G \tag{10} \quad G = \begin{pmatrix} p_1 & p_2 & \dots & p_m' \\ p_2 & p_2 & \dots & p_m' \\ \dots & \dots & \dots & \dots \\ p_m' & p_m' & \dots & p_m' \end{pmatrix} \quad p_i > p_{i+1}$$

更に、各列の列和から次の列の列和との差をとり n で割った値を対角成分とする対角行列 $D^{29)}$ を考えれば、この関係式はもっと簡単に、

$$G = T'DT \quad (11) \quad D = \begin{pmatrix} p_1 - p_2 & 0 & \cdots & 0 \\ 0 & p_2 - p_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_m \end{pmatrix}$$

と変形できる³⁰⁾。関係式(9), (10), (11)の計算は, バイナリデータ行列(M)より求めること



ができ, 行列 W の計算はこれらを利用して行うわけであるが, Horst の説明が不十分であるため行列 W の計算法を明確にするためには, コンピューターのプログラムによらざるを得ない. プログラムにおける BL 変数に前述した関係式のどの値が入っているかを見る為に, 図 2 に示したフローチャートによって計算の流れを追ってみた. 図 2 における I の部分は, 関係式(9)を利用して関係式(4)で示される列ベクトル DI を計算しているが, II の部分では, Horst が示した 8 つの関係式以外の 2 つの計算式(イ), (ロ)によって計算がなされている. 計算式(イ)における $VM(J)$ は各列の列和 (np) であり, $D(L)$ は列ベクトル DI である故に問題はないが, 計算式(ロ)における LL 行から LU 行までの和で示される $X(K, J)$, $J = 1, LL, K = LL, LU$ なる変数が何を意味するのか. (イ), (ロ)なる計算式が如何にして関係式(2), (3), (9), (10), (11)より導き出されたのか, こ

図 2 行列 W^{-1} のフローチャート

これらの点を説明することによって行列 W の計算法が明確になるはずである. まず, 計算式(ロ)で示される $X(K, J)$ なる変数がどのようなものであるかを図 3 に模式的に示してみた.

LL-LU	P(1)	NI×P(1)	D(1)	NI×D(1)	
8 - 10	1.0	10	0.3	3	
6 - 7	0.7	7	0.2	2	
3 - 5	0.5	5	0.3	3	
1 - 2	0.2	2	0.2	2	NL:4 NI:10

$$\begin{matrix}
 \begin{pmatrix}
 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{pmatrix} & \times & \begin{pmatrix}
 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 1 \\
 0 & 0 & 1 & 0 \\
 0 & 0 & 1 & 0 \\
 0 & 0 & 1 & 0 \\
 0 & 1 & 0 & 0 \\
 0 & 1 & 0 & 0 \\
 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0
 \end{pmatrix} & = & \begin{pmatrix}
 3 & 2 & 3 & 2 \\
 0 & 2 & 3 & 2 \\
 0 & 0 & 3 & 2 \\
 0 & 0 & 0 & 2
 \end{pmatrix}
 \end{matrix}$$

$M' \quad \times \quad A \quad = \quad X$

図3 行列Xと作業行列Aとの関係

図3によると、LL と LU はバイナリーデータ行列 (M) の列和によって決まり、バイナリーデータ行列 (M) における J 列の LL 行目から LU 行目までの各成分の和が X (K, J) である。X (K, J) を求めるといふことは、各列の列和の差を基準にして、図3に示したようにバイナリーデータ行列 (M) を(1)から(4)の小行列に分割し、それら小行列の列和を計算することである。このことを関係式で示すため A なる作業行列を考えれば、X(K, J) の計算を M'A で表わすことが可能である。但し、作業行列 (A) は、各列毎に LL 行から LU 行までの各成分のみが1.0を有し、他の成分は総て0.0という特殊な行列である。それ故、この行列の積和は関係式(11)における対角行列 (D) に等しい。つまり、

$$A'A = nD \tag{12}$$

になる。計算式(4)、(10)におけるプログラム上の変数 VM (J), D (L), X (K, J) は、それぞれ、nP, D1, M'A である故に計算式(4)、(10)は、次のような関係式として表現可能である。今、プログラム上の変数 BL (J, L) を行列Zとおくと

$$Z = M'A - nP (D1)'$$

になり、D は対角行列である故に関係式(9)より、

$$Z = M'A - (M'1) (1'D) \tag{13}$$

に変形できる。行列Z が Horst の関係式におけるどの行列に該当するかは、関係式(2), (3), (4), (9), (10), (11), (12)を利用して上記の関係式を展開することによって可能になるはずである。関係式(9), (10)より、

$$L'L = nG = nT'DT$$

の関係が成立し、n は整数である故に関係式(12)を利用すると

$$L'L = T'nDT = T'A'AT = (AT)'AT$$

になり

$$L = AT \tag{14}$$

という関係が成立する。行列 K は、関係式(2)より

$$K = nM'L - (M'I) (I'L)$$

で求まる故に関係式(14)を代入すると

$$K = nM'AT - (M'I) (I'AT)$$

になる。更に行列 T^{-1} を行列の右側からかけると

$$KT^{-1} = nM'ATT^{-1} - (M'I) (I'AT) T^{-1} = nM'A - (M'I) (I'A)$$

になり、関係式(3)より、

$$W = nM'A - (M'I) (I'A)$$

$$\frac{1}{n}W = M'A - \frac{1}{n} (M'I) (I'A) \tag{15}$$

になる。関係式(12)より、

$$\frac{1}{n}I'A = I'D$$

である故に関係式(15)は、

$$\frac{1}{n}W = M'A - (M'I) (I'D)$$

になり、関係式(13)に等しくなる。

$$Z = \frac{1}{n}W \tag{16}$$

となり、行列 Z は行列 W の $\frac{1}{n}$ 倍に等しい。つまり、プログラムでは Horst の関係式(2)によって行列 K をもとめるのではなく、作業行列 (A) を利用することによって関係式(13)を計算し、直接、行列 $\frac{1}{n}W$ を求めているのである。行列 H, F, nE, R は、それぞれ Horst の関係式(5), (6), (7), (8)を利用することによって求められる³¹⁾。

以上のような計算法によって残差相関行列 (R) を求めるわけであるが、LSSDM 法の特徴は、すでに指摘したように「バイナリーデータ行列の列和、或は行和はシンプレックスバイナリー行列のそれらに等しい。」という仮定にある。(但し、Horst の計算法では列和が等しい)³²⁾。そして、この仮定は行列 $\frac{1}{n}W$ の計算過程において、最も重要な関係として作用しているものであり、この仮定をもとに作業行列 (A) を利用して、仮説的なシンプレックスバイナリー行列 (L) とバイナリーデータ行列 (M) との積和を求めることも可能になるのである。Horst の(2), (3), (4)の関係式、また筆者が示した関係式(12), (13), (14), (15), (16)がこの解法を特徴付けていると言っても過言ではない。バイナリーデータ行列 (M) の列和、或は行和 (ここでは列和) が LSSDM 法に与える影響は非常に大きいと言えるのである。バイナリーデータ行列 (M) の列和において、 i 列と $i+1$ 列目の列和が等しい場合、 $np_i - np_{i+1} = 0$ になり、対角行列 (D) の i 列の対角成分が 0.0 になって行列 H の計算が不可能になる。つまり、各変数毎に計算される列和が等しい場合、或は非常に近似して列和の差が小さい場合、この解法は、使用不可能になるのである。その為、便宜的な計算法として、 $p_i = p_{i+1} - C$ ($C > 0$) なる定数 C を利用することが行われている³³⁾。しかし、定数 C は経験的に求めざるを得ず、また、この方法によって求まるシンプレックス

バイナリー行列 (L) の列和は、バイナリーデータ行列の列和とは等しくない。ここに、この解法の本質的な矛盾がある。そして、上記のような問題を有しているバイナリーデータ行列に関しては、様々な定数 C によって p_i を決定し 試行錯誤的に残差相関行列を求めていかざるを得ないのである。次に、この解法を適用した中心機能バイナリーデータ行列の因子分析における具体的な問題点を指摘したい。

2. 中心機能バイナリーデータ行列

(1) データと分析法

Berry は、行に中心集落を列に中心機能をとった Incidence Matrix において、各々の中心集落に中心機能が立地しているか否かを1.0, 0.0のバイナリーデータ行列として示した。筆者もこの Berry の方法にならい、60種の中心機能を変数にとり180の中心集落について中心機能の立地の有無を調べてみた³⁴⁾。その行列が行に中心集落、列に中心機能をとった180×60の中心機能バイナリーデータ行列である。これらのデータは、まず山口県全域について、昭和41年度の職業別電話帳をもとに180の中心集落に関して156の中心機能の有無を調べ、その中から中心集落の規模との相関が高い60種の中心機能を抽出したものである。更に、LSSDM 法によって残差相関行列を求める為、このデータ行列の行、列に関して次のような入替えを行った。行の上部から下部へ規模の大きな中心集落から規模の小さな中心集落がくるように行の並べ替えを行い、また、出現頻度の高い中心機能から出現頻度の低い中心機能が、この行列の左側から右側へ並ぶように列の並べ替えを行った³⁵⁾。

このようにして求められた中心機能バイナリーデータ行列から LSSDM 法によって残差相関行列を求め、この残差相関行列から共通因子の抽出を行ったのである。周知のように共通因子の抽出法には様々な解法があるが³⁶⁾、本研究では、共通因子の抽出と回転を同時に行う直接バリマックス法 (Direct Varimax Method)³⁷⁾ を利用した。これらの計算は、総て奈良女子大学計算機センターの FACOM270-20/30 の中型計算機によって行い、使用した総てのプログラムは、筆者が中型計算機用に修正し、組替えしたものである。直接バリマックス法のプログラムに関しては、芝裕順のプログラム³⁸⁾を参考に、最少二乗シンプレックスデータ行列解法のプログラムについては、Horst のそれを参考にした³⁹⁾。

(2) 作業行列 (A) と中心機能バイナリーデータ行列の列和との関係

表1は、中心機能バイナリーデータ行列の60変数(中心機能)に関する列和と p_i , d_i ⁴⁰⁾ を示したものである。この行列は、すでに行、列の並べ替えを行っている故に、60変数は出現頻度の高い変数から低い変数へと順次並んでいる。出現頻度が高いということは、数多くの中心集落に分散立地している低次の中心機能であることを示し、出現頻度が低いということは、少数の中心集落にのみ集中立地する高次の中心機能であることを示している。表1を見てもわかるように同一列和を有する中心機能が数多くあり、また出現頻度も低次中心機能から高次中心機能へと連続的に変化するのではなく、所々で不連続的に変化している。

中心機能バイナリーデータ行列に LSSDM 法を適用するわけであるが、関係式(4)で示した列ベクトル DI の各成分として $p_{i+1}-p_i=d_i$ とした場合、 d_{17} , d_{25} , d_{29} ……の13個の各成分は0.0値をとり、行列 H の計算が不可能になる。それ故、上記の関係式で d_i を求めるわけにはいかず、 $p_{i+1}-p_i=C$ なる関係式によって d_i を常に一定にし、シンプレックスバイナリー行列の p_i を決めることにした。問題は定数 C を如何にして決めるかということであるが、すでに指摘したように d_i を一定にした場合、理論的に最適な数値とい

表1 中心機能バイナリーデータ行列の列和と p_i, d_i

変数 番号	中心機能名	列和 (出現頻度)	p_i (列和/180)	d_i ($p_i - p_{i+1}$)	変数 番号	中心機能名	列和 (出現頻度)	p_i (列和/180)	d_i ($p_i - p_{i+1}$)
1	食品	177	0.9833	0.1556	31	洋品・雑貨	52	0.2889	0.0056
2	土木・建築・左官	149	0.8278	0.0333	32	電気工事	51	0.2833	0.0000
3	郵便局	143	0.7944	0.0333	33	フィルム・カメラ	51	0.2833	0.0000
4	医院	137	0.7611	0.0389	34	建築材料	51	0.2833	0.0333
5	荒物・日用雑貨	130	0.7222	0.0055	35	書籍	45	0.2500	0.0056
6	呉服・衣料	129	0.7167	0.0667	36	高校校	44	0.2400	0.0278
7	理容・美容	117	0.6500	0.0111	37	板金	39	0.2167	0.0222
8	燃料	115	0.6389	0.0444	38	家庭用具	35	0.1944	0.0055
9	野菜	107	0.5944	0.0055	39	清涼飲料水	34	0.1889	0.0111
10	旅館	106	0.5889	0.0667	40	印刷製本	32	0.1778	0.0000
11	食堂	94	0.5222	0.0111	41	娯楽場・ダンスホール	32	0.1778	0.0056
12	電気器具	92	0.5111	0.0111	42	靴	31	0.1772	0.0056
13	タクシ	90	0.5000	0.0722	43	鏡・ガラ	30	0.1667	0.0056
14	歯科	77	0.4278	0.0111	44	質屋	29	0.1611	0.0000
15	婦人子供服	75	0.4167	0.0167	45	映画・演芸	29	0.1611	0.0111
16	自転車部品	72	0.4000	0.0167	46	喫茶店	27	0.1500	0.0000
17	時計・めがね	69	0.3833	0.0000	47	ミシン・編物	27	0.1500	0.0056
18	新聞販売	69	0.3833	0.0055	48	通信部	26	0.1400	0.0000
19	金物	68	0.3778	0.0222	49	フトン・寝具	26	0.1400	0.0111
20	クリーニング	64	0.3556	0.0056	50	染色	24	0.1300	0.0056
21	文房具	63	0.3500	0.0222	51	生花	23	0.1278	0.0111
22	肉	59	0.3278	0.0056	52	印刷判	21	0.1167	0.0056
23	銀行	58	0.3222	0.0111	53	レコード・楽器	20	0.1111	0.0000
24	化粧品・小間物	56	0.3111	0.0055	54	旅行案内	20	0.1111	0.0000
25	履物	55	0.3056	0.0000	55	玩具	20	0.1111	0.0167
26	建具・表具	55	0.3056	0.0056	56	運動用具	17	0.0944	0.0111
27	建具・表具	54	0.3000	0.0056	57	各種学校	15	0.0833	0.0056
28	畳	53	0.2944	0.0056	58	装飾	14	0.0778	0.0056
29	あんま・はり・接骨	52	0.2889	0.0000	59	美術・骨とう	13	0.0722	0.0000
30	代書・司法書士	52	0.2889	0.0000	60	貸衣裳・貸フトン	13	0.0722	0.0722

注：コンピューターでは d_i の計算を列和の差/180.0 で計算したため、四捨五入の関係で $d_i = p_i - p_{i+1}$ にならない場合もある。

表 2 中心機能バイナリーデータ行列における作業行列 (A) と d_i との関係

変数番号	中心機能名	LL行~LU行	列和の差 ($VM_i - VM_{i+1}$)	d_i ($p_i - p_{i+1}$)	変数番号	中心機能名	LL行~LU行	列和の差 ($VM_i - VM_{i+1}$)	d_i ($p_i - p_{i+1}$)
1	食	150~177	28	0.1556	31	洋電	52~52	1	0.0056
2	土木・建築・左便	144~149	6	0.0333	32	品・雑貨	52~51	0	0.0000
3	郵便	138~143	6	0.0333	33	電気	52~51	0	0.0000
4	医	131~137	7	0.0389	54	イルム・カメ	46~51	6	0.0333
5	荒	130~130	1	0.0056	35	建	45~45	1	0.0056
6	兵	118~129	12	0.0667	36	書	40~44	5	0.0278
7	理	116~117	2	0.0111	37	高板	36~39	4	0.0222
8	燃	108~115	8	0.0444	38	家	35~35	1	0.0056
9	菓	107~107	1	0.0056	39	清	33~34	2	0.0111
10	旅	95~106	12	0.0667	40	涼飲	33~32	0	0.0000
11	食	93~94	2	0.0111	41	印	32~32	1	0.0056
12	電	91~92	2	0.0111	42	娯楽場・ダンスホール	31~31	1	0.0056
13	タ	78~90	13	0.0722	43	鏡・ガ	30~30	1	0.0056
14	歯	76~77	2	0.0111	44	質	30~29	0	0.0000
15	婦	73~75	3	0.0167	45	映	28~29	2	0.0111
16	バ	70~72	3	0.0167	46	喫	28~27	0	0.0000
17	自	70~69	0	0.0000	47	シ	27~27	1	0.0056
18	時	69~69	1	0.0056	48	通	27~26	0	0.0000
19	新	65~68	4	0.0222	49	フ	25~26	2	0.0111
20	金	64~64	1	0.0056	50	染	24~24	1	0.0056
21	ク	60~63	4	0.0222	51	生	22~23	2	0.0111
22	文	59~59	1	0.0056	52	印	21~21	1	0.0056
23	銀	57~58	2	0.0111	53	レ	21~20	0	0.0000
24	化粧品・小間物	56~56	1	0.0056	54	旅行	21~20	0	0.0000
25	履	55~55	1	0.0056	55	玩	18~20	3	0.0167
26	建	54~54	1	0.0056	56	運	16~17	2	0.0111
27	具・表	53~53	1	0.0056	57	各	15~15	1	0.0056
28	あんま・はり・接骨士	53~52	0	0.0000	58	種	14~14	1	0.0056
29	代	53~52	0	0.0000	59	術・骨と	14~13	0	0.0000
30	書	53~52	0	0.0000	60	美	1~13	13	0.0722

うものは明らかにされていない。試行錯誤的に定数Cを決定していく以外に方法はないのである。 d_i が最も重要な意味をもつのは、行列Wの計算においてであり、この場合、すでに指摘したように作業行列(A)の各成分が d_i によって決定される。そこで、作業行列(A)と d_i との関係を見ることによって定数Cを決めることにした。

表2は、 180×60 の大きさを有する作業行列(A)と d_i との関係を示したものである。LL行からLU行は、作業行列(A)において、1.0を有する行番号を各列毎(60変数について)に示したものである。第1列に関しては、150行目から177行目までが1.0を有し、他の成分は総て0.0であることを示している。つまり、180行のうち28の行だけが1.0を有する。この28の行は、第1列(第1変数)の列和から第2列(第2変数)の列和を引くことによって求まり、28を180行で割った値が第1列の $d_1=0.1556$ である。このように、作業行列(A)をバイナリーデータ行列と同一の大きさ(180×60)にした場合、列和の差は、作業行列(A)において各列毎に1.0を有する成分の数を示している。表2によると、作業行列(A)の各列が1つの行だけに1.0の成分を有する場合は、 d_i は0.0056になり、

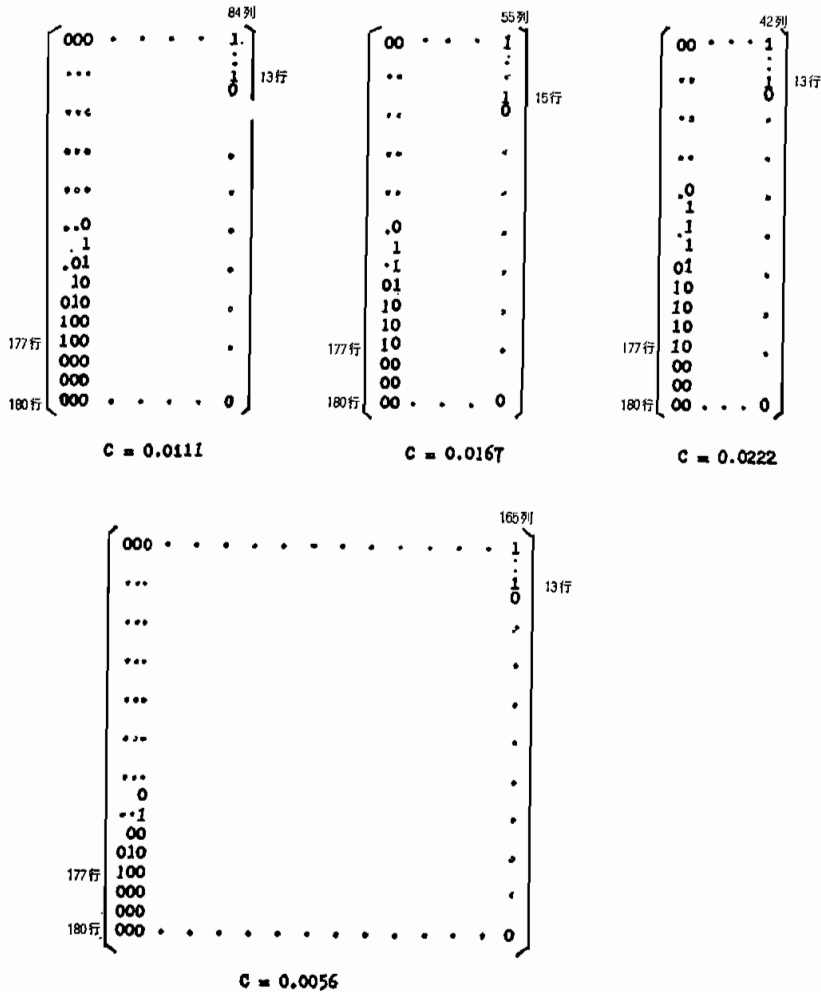


図4 Cと作業行列A($n \times m'$)との関係 $m' \leq m$

2つの行だけの成分が1.0の場合、 d_i は0.0111, 以下同様に3つの行だけの場合は、 d_i は0.0167, 4つの行だけの場合は、 d_i は0.0222……というように作業行列(A)の各成分と d_i との関係は、明白である。つまり、定数Cが0.0056, 0.0111, 0.0167, 0.0222……という値をとれば、作業行列(A)における総ての列に関して(但し、最終列は除く)1.0を有する成分の数が、図4に示したように1, 2, 3, 4, ……というように増加するわけである。作業行列(A)の大きさは、シンプレックスバイナリー行列の大きさに等しい。それ故、中心機能バイナリーデータ行列の場合も作業行列(A)は、行に180個の中心集落をとり、列に m' 個の変数をとることになる。この場合、 m' には $m' \leq m$, $m=60$ の制限があり、作業行列(A)の大きさと中心機能バイナリーデータ行列との大きさを等しくすることは、すでに指摘したように列和の差が0.0の変数が多くあるため不可能である。つまり、中心機能バイナリーデータ行列の場合、定数Cによって p_i を決定する故にバイナリーデータ行列とシンプレックス行列との大きさは、異ならざるを得ないのである。定数Cは、作業行列(A)の大きさにも規定されてくる故に、定数Cと作業行列(A)の大きさとの関係を調べ、

表3 定数Cと作業行列(A)の大きさ

定数 C	作業行列(A)の大きさ $n \times m'$
0.0056	180×165
0.0111	180×84
0.0167	180×55
0.0222	180×42
0.0278	180×33
0.0333	180×28
0.0389	180×24
0.0444	180×21

表3に示してみた。表3によると、定数Cが0.0056, 0.0111の場合、作業行列(A)の大きさは、180×165, 180×84となり列の数が制限の60を越えてしまう。それ故、定数Cとして妥当な値は、0.0167, 0.0222, 0.0278, 0.0333, 0.0389, 0.0444……であり、この場合、作業行列(A)の1.0を有する成分の数は、3, 4, 5, 6, 7, 8……となる。本研究では、定数Cとして0.0167, 0.0222, 0.0278, 0.0333, 0.0389, 0.0444の6つを取り上げ、各々の場合について、中心機能バイナリーデータ行列から残差相関行列を求め、それら各々について直接バリマックス

法を適用し、共通因子の抽出を行ってみた。しかし、これら総ての場合について、因子分析不可能という結果が出たのである。

(3) 中心機能バイナリーデータ行列の共通性における問題点

周知のように因子分析とは、共通性(communality)の推定値を対角成分とした相関行列から因子負荷行列を求め、共通因子構造を抽出しようとするものである。因子分析においては、この共通性の推定が非常に大きな意味をもっており⁴¹⁾、共通性の推定如何によって因子分析不可能という事態が生じる。共通性は、相関行列から求められる1.0よりも小さい正の実数であるが、ある種の相関行列ではこの共通性が1.0以上の値をとったり、負の値をとったりする場合がある。その場合、この相関行列から共通因子構造を抽出することはできない。勿論、このような現象は、共通性の推定方法に起因している場合もあり、相関行列そのものに起因する場合もある。本研究では、最も一般的な共通性の推定法であるSMC法(Squared Multiple Correlation Method)⁴²⁾を使用することにした。SMC法とは、重相関係数の平方(Squared Multiple Correlation)を用いる方法である。

表4は、前述した6種類の残差相関行列の共通性をSMC法で求めたものである。第1変数に関しては、総ての場合について共通性が負の値をとり⁴³⁾、更に、シンプレックスバイナリー行列の大きさ(列の数 m')が小さくなればなる程、他の変数にも負の値が出るようになる。このことは、総ての場合について因子分析が不可能なことを示しており、定数Cとして妥当な値は、0.0333以上であることを意味している。総ての場合について第1変数

表4 中心機能バイナリーデータ行列の共通性

変数 番号	中心機能名	DD(1) C=0.0167	DD(2) C=0.0222	DD(3) C=0.0278	DD(4) C=0.0333	DD(5) C=0.0389	DD(6) C=0.0444
1	食品	-1.358	-2.168	-1.117	-1.629	-1.642	-1.540
2	土木・建築・左官	0.128	0.143	0.119	0.109	0.059	0.015
3	郵便局	0.313	0.243	0.247	0.194	0.228	0.248
4	医院	0.146	0.108	0.048	0.063	-0.016	-0.044
5	荒物・日用雑貨	0.473	0.376	0.268	0.338	0.316	0.282
6	呉服・衣料	0.346	0.370	0.290	0.192	0.282	0.281
7	理容・美容	0.167	0.083	0.042	0.052	-0.088	0.015
8	燃料	0.468	0.252	0.251	0.255	0.259	0.201
9	薬局	0.335	0.253	0.197	0.205	0.184	0.190
10	旅館	0.261	0.289	0.276	0.221	0.204	0.227
11	食堂	0.457	0.438	0.323	0.377	0.380	0.345
12	電気器具	0.400	0.322	0.356	0.322	0.281	0.322
13	タクシ	0.513	0.396	0.384	0.355	0.352	0.327
14	歯科	0.349	0.331	0.407	0.317	0.269	0.313
15	婦人子供服	0.506	0.534	0.435	0.376	0.452	0.408
16	パ	0.542	0.540	0.510	0.468	0.496	0.450
17	自動車部品	0.547	0.514	0.552	0.472	0.468	0.456
18	時計・めがね	0.488	0.452	0.438	0.330	0.351	0.318
19	新聞販売	0.529	0.459	0.441	0.411	0.422	0.414
20	金物	0.539	0.460	0.453	0.423	0.442	0.459
21	クリーニング	0.649	0.515	0.475	0.412	0.445	0.400
22	文房具	0.539	0.517	0.463	0.463	0.415	0.414
23	肉	0.564	0.324	0.315	0.300	0.337	0.340
24	銀行	0.560	0.574	0.557	0.484	0.532	0.478
25	化粧品・小間物	0.672	0.571	0.562	0.510	0.504	0.515
26	履物	0.625	0.556	0.459	0.416	0.388	0.446
27	建具・表具	0.469	0.460	0.495	0.394	0.378	0.421
28	畳	0.567	0.547	0.543	0.491	0.511	0.510
29	あんま・はり・接骨	0.570	0.510	0.499	0.484	0.436	0.414
30	代書・司法書士	0.467	0.427	0.376	0.336	0.297	0.341
31	洋品・雑貨	0.534	0.509	0.519	0.500	0.403	0.399
32	電気工事	0.410	0.351	0.365	0.352	0.343	0.345
33	フィルム・カメラ	0.557	0.482	0.497	0.469	0.504	0.476
34	建材	0.682	0.621	0.593	0.586	0.599	0.534
35	書籍	0.588	0.616	0.547	0.456	0.445	0.485
36	高校	0.570	0.583	0.499	0.512	0.518	0.484
37	板金	0.594	0.558	0.507	0.499	0.531	0.488
38	家具	0.653	0.571	0.594	0.565	0.603	0.555
39	清涼飲料水	0.611	0.633	0.635	0.508	0.462	0.491
40	印刷・製本	0.678	0.595	0.589	0.534	0.556	0.524
41	娯楽場・ダンスホール	0.704	0.511	0.594	0.495	0.540	0.489
42	靴	0.790	0.736	0.703	0.686	0.680	0.659
43	鏡・ガラス	0.670	0.484	0.473	0.430	0.500	0.463
44	質屋	0.773	0.763	0.725	0.690	0.750	0.704
45	映画・演芸	0.569	0.538	0.531	0.507	0.496	0.503
46	喫茶店	0.730	0.677	0.686	0.714	0.642	0.584

表4 (続き)

変数 番号	中心機能名	DD(1)	DD(2)	DD(3)	DD(4)	DD(5)	DD(6)
		C=0.0167	C=0.0222	C=0.0278	C=0.0333	C=0.0389	C=0.0444
47	ミッション・編物	0.773	0.758	0.736	0.706	0.630	0.602
48	通信部	0.755	0.699	0.637	0.541	0.603	0.613
49	フットン・寝具	0.555	0.573	0.450	0.469	0.545	0.478
50	染色	0.593	0.658	0.531	0.501	0.580	0.559
51	生花	0.532	0.525	0.543	0.478	0.450	0.381
52	印刷	0.725	0.673	0.756	0.653	0.619	0.624
53	レコード・楽器	0.743	0.715	0.669	0.646	0.593	0.606
54	旅行案内	0.641	0.647	0.713	0.499	0.601	0.662
55	玩具	0.583	0.552	0.519	0.452	0.515	0.512
56	運動具	0.809	0.742	0.690	0.612	0.683	0.684
57	各種学校	2.004	0.569	0.701	2.001	0.501	0.633
58	装飾	0.490	0.482	0.464	0.404	0.478	0.396
59	美術・骨とう	0.524	0.478	0.526	0.455	0.450	0.462
60	貸衣裳・貸フトン	0.251	0.224	0.281	0.231	0.348	0.245

のみが負値をとるということは、第1変数の特殊性を意味するものである。SMC法において、共通性が負の値をとる場合は、相関行列の逆行列 (Inverse of Matrix) に問題があり、特にその対角成分が1.0以下であることを示している。表5は、各々の場合における残差相関行列の逆行列の対角成分を示したものである。この表によると、第1変数の対角成分は他の変数に比べて異常に低い。また、DD(5)以下の場合、負の共通性をとる第4、第7変数も対角成分が1.0に近い値をとるため、定数Cの影響によって共通性が負の値をとることもある。第57変数に関しては、常に共通性が異常なものではないが、DD(1)とDD(4)の場合、逆行列の対角成分が負の値をとり共通性が1.0以上の数値になる。この場合、第57変数の残差相関行列は、他の総ての変数との間に一の相関を示しており、この現象は、定数Cの数値による影響であると言える。結局、第1変数が負の共通性を有するのは、定数Cの数値に起因するのではなく、もとの中心機能バイナリーデータ行列に起因する現象であると言える。そこで、中心機能バイナリーデータ行列の共分散行列をみてみた。すると、第1変数と他の変数との共分散は、1.95から-0.71の範囲内にあるのに対し、他の変数間の共分散は、総て第1変数との共分散の10倍以上の数値を示しており、第1変数の共分散は異常に低いのである。このことは、共分散を求める関係式(1)において、第1変数の列和が他の変数のそれに比べて異常に大きいことに原因がある。つまり、第1変数の特殊性は、他の変数に比べて列和が大きすぎるということなのである。

LSSDM法においては、バイナリーデータ行列の列和が重要な意味を有していたことはすでに指摘したが、中心機能バイナリーデータ行列に関しては、変数間における列和にかなりの差がある為残差相関行列に影響を与え、因子分析を不可能にしているのである⁴⁾。

表5 残差相関行列の逆行列の対角成分

変数 番号	中心機能名	DD(3) C=0.0278	DD(4) C=0.0333	DD(5) C=0.0389	変数 番号	中心機能名	DD(3) C=0.0278	DD(4) C=0.0333	DD(5) C=0.0389
1	食品	0.4723	0.3803	0.3785	31	洋品・雑貨	2.0801	1.9988	1.6759
2	土木・建築・左官	1.1353	1.1225	1.0630	32	電気工事	1.5751	1.5438	1.5225
3	郵便局	1.3286	1.2400	1.2959	33	フィルム・カメラ	1.9869	1.8806	2.0177
4	医物・日用雑貨	1.0505	1.0674	0.9835	34	建築材料	2.4596	2.4152	2.4913
5	荒服・衣料	1.3668	1.5111	1.4623	35	書籍	2.2066	1.8373	1.8012
6	呉服・衣料	1.4085	1.2378	1.3927	36	高校	1.9972	2.0486	2.0763
7	理容・美容	1.0447	1.0053	0.9193	37	板金	2.0275	1.9950	2.1328
8	燃料	1.3351	1.3432	1.3503	38	家具	2.4638	2.2998	2.5171
9	薬局	1.2451	1.2583	1.2257	39	清涼飲料水	2.7387	2.0326	1.8581
10	旅館	1.3807	1.2838	1.2558	40	印刷・製本	2.4313	2.1609	2.2511
11	食堂	1.4766	1.6050	1.6131	41	娯楽場・ダンスホール	2.4630	1.9803	2.1740
12	電気器具	1.5531	1.4979	1.3910	42	靴	3.3720	3.1802	3.1261
13	タシ	1.6222	1.5514	1.5440	43	鏡・ガラス	1.8974	1.7557	2.0019
14	歯科	1.6852	1.4649	1.3685	44	質屋	3.6330	3.2274	4.0071
15	婦人子供服	1.7709	1.6036	1.8260	45	映画・演芸	2.1308	2.0298	1.9856
16	自動車部品	2.0401	1.8803	1.9855	46	喫茶店	3.1868	3.4994	2.7898
17	時計・めがね	2.2338	1.8950	1.8808	47	ミシン・編物	3.7815	3.3966	2.7035
18	新聞販売	1.7807	1.4920	1.5408	48	通信部	2.7525	2.1804	2.5185
19	金物	1.7896	1.6981	1.7288	49	フトン・寝具	1.8170	1.8815	2.1980
20	クリーニング	1.8284	1.7339	1.7910	50	染色	2.1301	2.0026	2.3794
21	クリーニング	1.9054	1.7011	1.8014	51	生花	2.1879	1.9159	1.8168
22	文房具	1.8632	1.8634	1.7101	52	印刷判	4.1065	2.8850	2.6277
23	肉	1.4608	1.4289	1.5094	53	レコード・楽器	3.0156	2.8241	2.4567
24	銀行	2.2562	1.9371	2.1361	54	旅行案内	3.4864	1.9974	2.5053
25	化粧品・小間物	2.2865	2.0411	2.0165	55	玩具	2.0804	1.8247	2.0627
26	履物	1.8490	1.7118	1.6341	56	運動具	3.2244	2.5783	3.1581
27	建具・表具	1.9792	1.6497	1.6065	57	各種学校	3.3492	-0.9991	2.0040
28	畳	2.1858	1.9637	2.0451	58	装飾	1.8643	1.8324	1.9172
29	あんま・はり・接骨	1.9948	1.9390	1.7735	59	美術・骨とう	2.1077	1.8340	1.8195
30	代書・司法書士	1.6036	1.5062	1.4222	60	貸衣裳・貸フトン	1.3915	1.3017	1.5345

確井：中心機能バインナーデータ行列の因子分析における問題点

3. 結 語

60種の中心機能を変数とした中心機能バイナリーデータ行列において、この行列の列和は、各々の中心機能がどれだけの中心集落に保有されているかを示している。つまり、列和が大きくなればなる程多くの中心集落にその機能が分散立地していることを示し、列和が等しいということは、それらの中心機能がよく似た立地傾向を示すということの意味しているのである。更に、この行列において隣接する各変数間の列和の差が均一でなく、分散が大きいうことは、列和の大きい分散立地型の中心機能から列和の小さい集中立地型の中心機能へと中心機能の立地が、連続的に変化しているのではなく、段階的につまり不連続的に変化していることを示している。そしてこれらのことは、中心機能の階層的特性を暗示しているわけであるが、この中心機能の階層的特性が LSSDM 法の適用を不可能にしているのである。

最少二乗シンプレックスデータ行列解法は、バイナリーデータ行列の変数間の相関を求めるため、変数そのものに起因するみかけの相関（項目の選好性や難易度など）をシンプレックスバイナリー行列という形で除去し、真の相関係数を求めようとするものである。そしてこのシンプレックスバイナリー行列の特性を「バイナリーデータ行列の列和、或は行和は、シンプレックスバイナリー行列のそれらに等しい。」と規定したところに、この解法の特徴がある。筆者が明らかにしたように総ての計算は、バイナリーデータ行列の列和を基にして行われ、特に行列 $\frac{1}{n}W$ の計算過程において作業行列 (A) が、バイナリーデータ行列の列和に規定されていることはすでに指摘した通りである。この解法において、上記の仮定を満足させる為には、列和の等しい変数を有しないことがバイナリーデータ行列の暗黙の条件となっている。しかし、Horst は、この解法のバイナリーデータ行列の条件として、特にこのことには言及していないのである。それ故、中心機能バイナリーデータ行列のように列和の等しい変数が幾つかある場合や、列和の差が非常に小さい変数である場合など、便宜的に $p_{i+1}-p_i=C$ なる関係式でシンプレックスバイナリー行列の列和を決めることが一般的に行われている。しかし、この場合、バイナリーデータ行列とシンプレックスバイナリー行列との列和は等しくなく、前述の仮定とは矛盾しているのである。更に、この関係式によって定数Cでシンプレックスバイナリー行列の列和を決めることは、列和の差が均一であることを意味し、中心機能バイナリーデータ行列のように変数間の列和の差が均一でない場合、もとのバイナリーデータ行列とはかけ離れたシンプレックスバイナリー行列を求めていることになる。つまり、この場合もシンプレックスバイナリー行列とバイナリーデータ行列との近似性というこの解法の原理に矛盾してしまうのである。

結局、LSSDM 法は、各変数の列和が比較的均一で、同一列和を有しない変数のみからなるバイナリーデータ行列の場合のみ有効であるが、それ以外の例えば中心機能バイナリーデータ行列のような場合には、適用しない方がよいと言える。

中心機能の分類において、中心機能バイナリーデータ行列のようなバイナリー型のデータを利用することは、中心機能の分散的、或は集中的立地傾向から分類を試みるという点において意義はあっても、適当な分析法が見付からない以上、この種のデータによる中心機能の客観的な分類は不可能である。もともと因子分析がバイナリーデータには不適な方法である故に、因子分析を利用するならばバイナリーデータ型以外のデータ行列を求めた方がよい。

本稿の作成にあたり、資料の提供をはじめ、終始御指導を賜った奈良女子大学の西村

睦男先生、プログラムの作成にあたり貴重な御助言をいただいた奈良女子大学、物理学科の村井友和先生に厚く御礼申し上げます。また、計算機の使用をはじめ、プログラム作成の御指導をいただいた奈良女子大学計算機センターの高津秀子技官、足立真知子さんに感謝致します。

註

1. Brush, J.E.: "The Hierarchy of Central Places in Southwestern Wisconsin": Geogr. Rev. vol. 43 pp. 380~402, 1953.
渡辺良雄：“東北地方における中心地の階層分化” 東北地理 vol. 19 No. 1 pp. 1~9, 1967.
森川 洋：“勢力圏設定に関する一考察—岡山県を例として—” 人文地理 vol. 19 No. 1 pp. 31~53, 1967.
これ以外に数多くの文献があげられるが、古典的な方法の代表として Brush の論文と日本独自の方法である渡辺、森川の論文をあげておく。
2. Berry, B. J. L. and Barnum, H.G.: "Aggregate relations and elemental components of central places system" Journal of Regional Science vol. 4 No. 1 pp. 35~68, 1962.
3. Horst, P.: "Factor analysis of data matrices." Holt, Rinehart & Winston. Chapter 12. pp. 258~283, 1965.
4. 芝 祐順：“行動科学における相関分析法” 東京大学出版会, 1972.
5. Berry, *ibid.* p. 46.
6. *Ibid.* pp. 42~46.
7. *Ibid.*
8. Horst, *ibid.* pp. 518~532.
9. *Ibid.* pp. 696~697.
10. *Ibid.* p. 513.
11. *Ibid.* p. 514.
12. *Ibid.* pp. 514~515.
13. *Ibid.* p. 514.
14. 安田三郎：“社会統計学” 九善. pp. 44~46, 1969.
15. Horst, *ibid.* p. 514.
16. *Ibid.* p. 514.
17. 一般的に残差相関行列といえば、相関行列から相関行列を引いたものをいうが、この場合は、残差バイナリー行列の相関行列ということである。
18. これ以外の解法として、最少二乗シンプレックス共分散行列解法 (The least squared simplex covariance matrix solution) やショートカットシンプレックス共分散行列解法 (Computational short cut for the simplex covariance matrix solution) などがある。
19. Horst, *ibid.* p. 517.
20. 独立変数 x は、バイナリーデータ行列 (M) の偏差型、つまり、 $x = \left(I - \frac{11'}{u} \right) M$ に等しく、従属変数 y は、シンプレックスバイナリー行列 (L) の偏差型、 $y = \left(I - \frac{11'}{n} \right) L$ に等しい。
21. x の係数 B は、関係式 $B = (x'x)^{-1}x'y$ で求められる。
22. Horst, *ibid.* pp. 527~532.
23. 安田三郎：前掲書. pp. 171~186.
24. Horst, *ibid.* pp. 518~519.
この8つの式のうち(5), (6), (7)において、Horst は、 $n^2H=H$ $n^2F=F$ $nE=E$ と置き換えているが、誤解をさけるためもとの形で示した。
25. Horst は、バイナリーデータ行列の列和を nP 、シンプレックスバイナリー行列の列和を $n\theta$

として区別しているが、実質的には同一のものであるので本論文では総て nP として示した。なお、関係式の中で使用されている I は、単位ベクトルである。

26. Horst, *ibid.* p. 518.
27. *Ibid.* p. 527.
28. *Ibid.* p. 529.
29. 最終列の対角成分は例外で最終列の列和に等しい。
30. Horst, *ibid.* p. 530.
31. これらの行列の計算は、関係式の通りで問題はない。
32. 列和か行和かのどちらかが等しければよく、とり方によって列和か行和かになる。
33. Horst, *ibid.* p. 532.
34. このデータは、西村隆男教授の調査資料による。
35. 行、列の並べ替えに関するコンピュータープログラムは、筆者がこのデータ用に作成した。
36. 芝 裕順：因子分析法 東京大学出版会、1972.
37. 前掲書 pp. 117~140.
38. 同上 pp. 383~384.
39. 使用したプログラムは、論文の最後に掲載した。
40. p_i は列ベクトル P の各成分、 d_i は列ベクトル DI の各成分を示す。
41. 芝裕順；前掲書 (36) pp. 161~191.
42. 同上 pp. 172~178, pp. 390~391.
43. 共通性の指定法として、SMC 法以外に相関係数の和の平方と全相関係数の比をもって推定する方法 $(h_j^2 = (\sum_{k=1}^I r_{jk})^2 / \sum_{k=1}^I \sum_{h=1}^I r_{jh})$ も利用したが、この場合は、残差相関行列に負の相関が多い故に、総ての変数について負の共通性を示した。変数が多い場合、最も理想的な共通性の指定法は SMC 法だと言われており、バイナリーデータ行列の場合も他の方法に比べてこの方法が最もよいと思われる。
44. 第1変数を除いた中心機能バイナリーデータ行列の場合も第4変数(医院)の共通性はすべての場合に負値をとり、因子分析は不可能であった。列和が大きくなるほど共通性は小さくなり、極端な場合は負値になる傾向がある。

Summary

According to order of size-rank in central places, central functions are higher in level. It is generally understood that a hierarchical structure is controlled by order of central functions. Many articles have concerned with the classification of central functions in order to analyze a hierarchical structure in central place system, but those methods have been obscure and nonscientific.

Perhaps the work of Berry and Burnum is the first article classified central functions by a scientific method; a factor analysis. The method of a factor analysis that they used is the direct factor analysis, but we cannot use this method, because it is a special kind of a factor analytic solution and its computer program is not present in their paper.

So another factor analytic solutions which we have used are the least squared simplex binary data matrix solution and the direct varimax factoring solution.

Unfortunately, we have failed to solve the central function-binary data matrix by those solutions, especially by the least squared simplex binary data matrix solution. Therefore it is the purpose of this paper to point out why the least squared simplex

binary data solution could not be applied to the central function–binary data matrix.

The central function–binary data matrix is the data matrix in which each element is either 1 or 0, of which the rows are central places, the columns are central functions (variables).

In a binary data matrix, the fact that correlations among variables are influenced by the dispersion of item preference and difficulty has been known. We must remove this phenomenon which is due to variation in item preference and difficulty from a binary data matrix. In general, some of mathematical solutions have been used to remove this phenomenon. Those solutions are based on the notion of the simplex which Guttman developed. The general approach is to apply a least squared transformation to this binary data matrix. From here we proceed to find a residual matrix, and then a residual correlation matrix. This residual correlation matrix is freed from the effects of variation in item preference and difficulty. This solution is called the least squared simplex binary data matrix solution and it must satisfy the restriction that the row sums or the column sums of a simplex binary matrix will be the same as for a binary data matrix.

This restriction is due to the unsuccess of the analysis in the central function–binary data matrix. In this matrix, the column sums of some variables are equal.

But, when the column sums of a binary data matrix may be equal, it is impossible to apply this solution to a binary data matrix. In this case another procedure for determining the column sums of a simplex binary matrix is to use the equation, which is indicated by $p_i - p_{i+1} = C$, p_i is elements of m times a column vector of the column sums in a simplex binary matrix, m is the number of variables.

By this equation, we determined the column sums of the simplex binary matrix in the central function–binary data matrix, then proceeded to find the residual correlation matrix of this matrix. However, the residual correlation matrix had some negative communalities which led us into more difficulty and we could not apply the direct varimax factoring solution to this residual correlation matrix.

We consider that the cause of negative communalities must be the column sums of the central function–binary data matrix. Because, though p_i in this equation is equally spaced, p_i in the central function–binary data matrix is not equally spaced.

At last, because of the great deviation of the column sums of the central function–binary data matrix, it is impossible to apply the least squared simplex binary data matrix solution to this central function–binary data matrix.

附表：最少二乗シンプレックスバイナリーデータ行列解法のコンピュータープログラム

FACOM 270-20/30 FORTRAN LIST

```

C CORRELATION OF RESIDUAL MATRIX IN BINARY DATA
COMMON BB(60,60),VM(60),P(61),D(60),S(60),SQ,N,NS,NP,LL,LU,I,J,L
DRUM DIMENSION DX(180,60),DRL(60,60)
SEG ENTRY BCOV,DP,BLCOV,HMATRX,FMATRX,EFCURR,INP,OUTP
CALL INP(1)
CALL OUTP(1)
CALL BCOV(N,NS)
CALL OUTP(2)
10 CALL DP(NP)
CALL OUTP(3)
P(NP+1)=0.0
FNS=NS
DO 153 L=1,NP
LL=FNS*P(L+1)+1.1
LU=FNS*P(L)+0.1
CALL OUTP(4)
DO 153 J=1,N
CALL BLCOV(LL,LU,J,L)
153 CONTINUE
CALL OUTP(6)
CALL HMATRX(N,NP)
CALL OUTP(7)
CALL FMATRX
CALL OUTP(8)
CALL INP(2)
DO 166 I=1,N
CALL INP(3)
DO 165 J=1,N
165 BB(I,J)=VM(J)-BB(I,J)
166 D(I)=1.0/SQRT(BB(I,I))
CALL OUTP(9)
CALL EFCURR(N)
CALL OUTP(10)
CALL INP(4)
CALL OUTP(1)
GO TO 10
STOP
END

SUBROUTINE BCOV(N,NS)
COMMON BB(60,60),VM(60)
DRUM DIMENSION DX(180,60)
DIMENSION XI(180),XJ(180)
FNS=NS
DO 142 I=1,N
DRUM READ DX(I,I),(XI(K),K=1,NS)
DO 142 J=1,N
DRUM READ DX(I,J),(XJ(K),K=1,NS)
BB(I,J)=-VM(I)*VM(J)/FNS
DO 142 KK=1,NS
BB(I,J)=BB(I,J)+XI(KK)*XJ(KK)
142 CONTINUE
RETURN
END

SUBROUTINE DP(NP)
COMMON BB(60,60),VM(60),P(61),D(60)
NP1=NP-1
DO 144 I=1,NP1
144 D(I)=P(I)-P(I+1)
D(NP)=P(NP)
RETURN
END

SUBROUTINE BLCOV(LL,LU,J,L)
COMMON BL(60,60),VM(60),P(61),D(60),S(60),SQ,N,NS,NP
DIMENSION XI(180)
DRUM DIMENSION DX(180,60)
DRUM READ DX(I,J),(XI(JJ),JJ=1,NS)

```

附表；（続き）

```

BL(J,L)=0.0
BL(J,L)=-VM(J)*D(L)
DO 154 K=LL,LU
BL(J,L)=BL(J,L)+XI(K)
154 CONTINUE
RETURN
END

SUBROUTINE HMATRX(N,NP)
COMMON H(60,60),VM(60),P(61),D(60)
DRUM DIMENSION DX(180,60),DBL(60,60)
DIMENSION BI(60),BJ(60)
DO 162 I=1,N
DRUM READ DBL(1,I),(BI(KK),KK=1,N)
DO 162 J=1,N
H(I,J)=0.0
DRUM READ DBL(1,J),(BJ(KK),KK=1,N)
DO 162 K=1,NP
162 H(I,J)=H(I,J)+BI(K)*BJ(K)/D(K)
RETURN
END

SUBROUTINE FMATRX
COMMON F(60,60),VM(60),P(61),D(60),S(60),SQ,N,NS,NP
DRUM DIMENSION DX(180,60),DBL(60,60)
DIMENSION XI(180)
FNS=NS
NP1=FNS*P(1)+0.1
DO 1626 J=1,N
S(J)=-VM(J)*P(1)
DRUM READ DX(1,J),(XI(K),K=1,NS)
DO 1626 I=1,NP1
S(J)=S(J)+XI(I)
1626 CONTINUE
SQ=1.0/SQRT(1.0-P(1))
DO 1629 J=1,N
1629 S(J)=S(J)*SQ
DO 1632 I=1,N
DO 1632 J=1,N
1632 F(I,J)=F(I,J)/FNS+S(I)*S(J)/FNS
RETURN
END

SUBROUTINE EECORR(N)
COMMON RE(60,60),VM(60),P(61),D(60)
DO 169 I=1,N
DO 169 J=1,N
169 RE(I,J)=D(I)*RE(I,J)*D(J)
RETURN
END

SUBROUTINE INP(KK)
COMMON BB(60,60),VM(60),P(61),D(60),S(60),SQ,N,NS,NP,LL,LU,LI
DRUM DIMENSION DX(180,60),DBL(60,60)
DIMENSION XI(180)
GO TO (1,2,3,4)KK
1 CONTINUE
READ(4,903)N,NS,NP
DO 10 I=1,N
VM(I)=0.0
P(I)=0.0
D(I)=0.0
S(I)=0.0
DO 10 J=1,N
BB(I,J)=0.0
10 CONTINUE
READ(4,905)(P(I),I=1,NP)
READ(4,905)(VM(I),I=1,N)
DO 11 I=1,N
READ(4,904)(XI(J),J=1,NS)
DRUM WRITE DX(1,I),(XI(J),J=1,NS)
11 CONTINUE

```

附表：(続き)

```

RETURN
2 CONTINUE
DO 12 J=1,N
READ(4,906)(VM(I),I=1,N)
12 DRUM WRITE DBL(1,J),(VM(I),I=1,N)
RETURN
3 CONTINUE
DRUM READ DBL(1,11),(VM(K),K=1,N)
RETURN
4 CONTINUE
READ(4,903)N,NS,NP
READ(4,905)(P(I),I=1,NP)
READ(4,905)(VM(I),I=1,N)
RETURN
903 FORMAT(3I4)
904 FORMAT(20F3.0)
905 FORMAT(8E15,5)
906 FORMAT(8E15,7)
END

SUBROUTINE OUTP(KKK)
COMMON BB(60,60),VM(60),P(61),D(60),S(60),SQ,N,NS,NP,LL,LU,
111,J1,L1
DRUM DIMENSION DX(180,60),DBL(60,60)
DIMENSION X1(180)
GO TO (1,2,3,4,5,6,7,8,9,10)KKK
1 CONTINUE
WRITE(6,700)
WRITE(6,703)N,NS,NP
WRITE(6,705)(P(I),I=1,NP)
WRITE(6,706)(VM(I),I=1,N)
RETURN
2 CONTINUE
WRITE(6,800)
DO 1422 I=1,N
WRITE(6,904)
WRITE(6,906)(BB(I,J),J=1,N)
1422 CONTINUE
DO 1423 I=1,N
1423 WRITE(8,905)(BB(I,J),J=1,N)
RETURN
3 CONTINUE
WRITE(6,811)
WRITE(6,906)(D(I),I=1,NP)
RETURN
4 CONTINUE
WRITE(6,802)LL,LU,D(L1)
RETURN
5 CONTINUE
WRITE(6,803)VM(J1),D(L1),BB(J1,L1)
RETURN
6 CONTINUE
WRITE(6,806)
DO 1572 I=1,N
WRITE(6,904)
WRITE(6,906)(BB(I,J),J=1,NP)
DRUM WRITE DBL(1,I),(BB(I,J),J=1,NP)
1572 CONTINUE
RETURN
7 CONTINUE
WRITE(6,908)
DO 801 I=1,N
WRITE(6,904)
801 WRITE(6,906)(BB(I,J),J=1,N)
RETURN
8 CONTINUE
WRITE(6,907)SQ
WRITE(6,911)
WRITE(6,906)(S(J),J=1,N)
WRITE(6,913)

```

附表：(続き)

```

DO 1634 I=1,N
WRITE(6,904)
1634 WRITE(6,906)(BB(I,J),J=1,N)
RETURN
9 CONTINUE
WRITE(6,914)
DO 167 I=1,N
WRITE(6,904)
167 WRITE(6,919)(BB(I,J),J=1,N)
WRITE(6,910)(D(I),I=1,N)
RETURN
10 CONTINUE
WRITE(6,916)
DO 171 I=1,N
WRITE(6,904)
171 WRITE(6,919)(BB(I,J),J=1,N)
DO 172 I=1,N
172 WRITE(8,905)(BB(I,J),J=1,I)
RETURN
700 FORMAT(1H0,34HRESIDUAL COVARIANCE OF BINARY DATA,5X,
119HPROGRAM BY HORST P.,5X,16HMODIFIED BY USUI)
703 FORMAT(1H1,3X,2HN=,I4,5X,3HNS=,I4,5X,3HNP=,I4)
705 FORMAT(1H0,3X,2HP=,/,8X,10F7.4)
706 FORMAT(1H0,3X,15HVM=SUM OF COLUM,/,8X,10F7.0)
800 FORMAT(1H1,3X,35HBB=COVARIANCE MATRIX OF BINARY DATA)
904 FORMAT(1H0)
906 FORMAT(5F15.4)
905 FORMAT(8E15.7)
811 FORMAT(1H1,3X,16HD(I)=P(I)-P(I+1),5X,2HD=)
802 FORMAT(1H0,3X,3HLL=,I4,5X,3HLU=,I4,5X,2HD=,F10.4)
803 FORMAT(1H0,3X,3HVM=,F8.0,5X,2HD=,F10.4,5X,3HBL=,F10.4)
806 FORMAT(1H1,3X,63HBL=COVARIANCE MATRIX OF BINARY DATA BETWEEN SIMPL
1EX BINARY DATA)
908 FORMAT(1H1,3X,10HH= HMATRIX)
907 FORMAT(1H1,3X,3HSQ=,F10.4)
911 FORMAT(1H0,3X,12HS(J)=S(J)*SQ)
913 FORMAT(1H1,3X,10HF= FMATRIX)
914 FORMAT(1H1,3X,18HRE=RESIDUAL MATRIX)
916 FORMAT(1H1,3X,30HCORRELATION OF RESIDUAL MATRIX)
910 FORMAT(1H0,3X,2HD=,(5F10.4))
919 FORMAT(10F8.4)
END

```