

# 因子分析法のための会話型プログラム

教育心理学研究室 芝

祐 順

## A Conversational Program for Factor Analysis

Sukeyori SHIBA

The purpose of this study is to develop a conversational factor analysis program for computer use. Process of factor analysis contains a series of computational procedures. Factor analysts usually compute several different solutions to explore the best one in reference to the purpose of his study. This program provides many options on preparatory data processing, for example, data correction, data transformation and finding outliers, initial factor solutions, estimation of the number of factors, rotation of factor axis and estimation of factor scores. An investigator is allowed to examine various solutions in a variety of conditions without any special background in computer usage.

In addition, the program can be used as a kind of computer assisted instruction on factor analysis which gives students step by step exercises. Students will be able to master the process of factor analysis in responding the conversational program without conducting complicated computational works.

因子分析は多変量データの中に潜む相関的構造を解析するのに有効であるが、その利用方法の様式は、事前の情報の多寡によって違いがある。それまでに十分解析結果が得られているような変数を主とする場合には、因子構造について一定の情報を仮說的にもっており、新しいデータの因子構造も、これに対応づけて検討されていることが多い。これに対し、それまでに関連する解析結果がないときには、変数間の相関的関係も十分明らかでないため、特定の構造を予想して解析することはできない。このような状況では、因子分析の過程で得られた結果を即座に利用して、再分析をすすめるという柔軟な手続きが有効である。

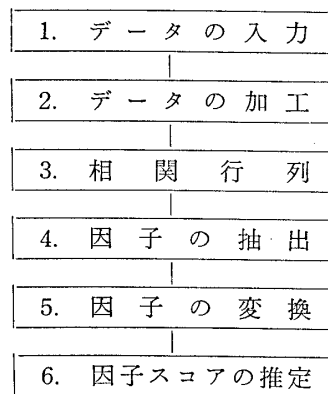
因子分析に関しては、これまで多数の計算プログラムが開発され、その利用も容易になった。しかし、それらはいずれもバッチ処理用のものであるため、計算結果をみながら再分析を続けるというやり方には適していない。ここに述べる会話型因子分析プログラムは、因子分析の基本的な解析過程を自由に選択しながら連続的に実行可能にするものである。これにより、従来かなり手間のかかった再分析を、極めて容易に実行できるようになった。

このプログラムを開発したもう一つの動機は、因子分析法の演習のための利用にある。そのため、プログラム中には実用的観点からすれば、必ずしも必要ではない解法についても、オプションの中に加えてある。

### I

#### A. プログラムの構成

プログラムは、図のように6つのコンポーネントから成っている。



それぞれのコンポーネントには、利用者が選択できる

オプションが用意されている。その概要は次の通りである。

#### 1-1 データの入力方法

- ① キーボードより直接入力する。
- ② 既に格納済みのデータセットより入力する。

#### 1-2 入力できるデータの種類

- ① 観測値行列
- ② 相関行列
- ③ 因子負荷行列

#### 2-1 因子分析に入る前の、入力データの加工、調整

- ① 変数、個体、因子などの削除
- ② タイトル名、変数名、個体名、因子名、データの値などの修正
- ③ 外れ値の検出
- ④ データ変換

#### 3-1 相関行列をもとめるときのモデル

- ① 変量間の相関 (R-技法)
- ② 個体間の相関 (Q-技法)

#### 3-2 標準化

- ① R-技法の場合、相関行列の計算の前に、個体単位毎の標準化をおこなうことができる。
- ② Q-技法の場合、変量単位毎の標準化をおこなうことができる。

#### 3-3 相関係数の種類

- ① 積率相関係数
- ② 四分相関係数

#### 4-1 因子抽出法

- ① 主因子法 (乗ベキ法によるもの)
- ② 主因子法 (ヤコビ法によるもの)
- ③ 最小残差法
- ④ セントロイド法
- ⑤ 最尤法

#### 4-2 共通性の推定

- ① SMC
- ② 非対角要素の最大値
- ③ 1.0
- ④ キーボードから入力

#### 4-3 因子数の決定方法

- ① キーボードから入力
- ② 簡便法
- ③ 統計的仮説検定による方法
- ④ 最尤法の場合、AIC による方法

#### 5-1 因子の変換

- ① バリマックス法
- ② オーソマックス法

③ ジェオマックス法 (直交)

④ ジェオマックス法 (斜交)

⑤ 一般オブリンミン法

⑥ プロクラステス法 (直交)

⑦ プロクラステス法 (斜交)

#### 6-1 因子スコアの推定

- ① 推定式  $F_7$  によるもの
- ② 推定式  $F_{13}$  によるもの

### B. プログラム利用上の一般的注意

#### 1. 利用機種

東京大学大型計算センターの HITAC M-200H システム (VOS 3 システム) を、ディスプレイ端末装置から使用する。使用言語は FORTRAN 77 (一部 VOS 3 最適化 FORTRAN 利用)。使用容量は、ソースデータセットは 600 KB、ロードモジュールセットはそれより多少大きくなる。

#### 2. 呼出し方

VOS 3 における端末からの呼び出し方は次の通りである。

- ① キーボードから、データを入力する場合

```
»FACTA
```

- ② 格納済みのデータ・セットから入力する場合

```
»FACTA△DS(ABC. DATA)
```

ABC・DATA はデータが格納されているデータ・セット名である (4 項の補足 1 参照)

#### 3. 入力データの制限について

現在の大型センターのジョブクラス制限の標準値のもとでは、個体数は最大 300、変量数は最大 30、因子数は最大 30 である。しかし、制限の上限をあげれば更に拡張できる。その際には、メインプログラムのパラメータ文を変更する必要がある。

#### 4. データの入力方法

データの入力方法には、プログラムの実行段階で、キーボードから直接入力する方法と、あらかじめ入力データのデータセットを作って格納しておく方法の 2 通りがある。

- ① キーボードから直接入力する方法

比較的少量のデータの場合は、プログラムの実行段階で、ディスプレイからの指示に従って、直接入力する方法が便利である。

なお、今後入力データとして再利用する場合は、システム側で用意したデータセットに出力し、保存することができる。

```

TYPE DATA(I,J) WITHIN 8 DIGITS(EX. 83.4)
DATA( 1, 1)...?
15.
DATA( 1, 2)...?
55.5
    
```

② あらかじめ、入力データのデータセットを作っておく方法

多量のデータを用いる場合は、あらかじめ入力データのデータセットを作成しておくことと便利である。この場合、カード穿孔してカードリーダーから入力する方法と、TSS の EDIT コマンドを用いてキーボードから入力する2通りの方法がある。

データの書式には、標準書式と標準外書式の2種類がある。標準書式とは、入力データのほかにデータタイトル及び入力データの大きさ等を付加するもので、プログラム実行段階の応答手続が簡略化される利点がある。これに対し、標準外書式は入力データの行列のみから成っている。標準外書式のデータもプログラム実行中に出力ファイルオプションを利用すれば、標準書式に変換した新たなデータセットになる。

例1 カードリーダーより、標準フォーマットのデータカードを入力させる場合

NO.	5	10	15	20	25
1	//A1234XYZ JOB パスワード, CLASS=A				
2	>>USE				
3	>>DATA ,ABC.DATA				
4	01				
5	タイク ソクテイ データ (1981-04-10)				
6	25.5				
7	6				
8	1.4	170.3	39.45	16.73	
9	0.05	162.4	37.0	12.1	
10					255組
11					
12	01				
13	タイク イ知ウ				
14	イソク ロクハ				
15					255枚
16					
17	01				
18	タイショウ				
19	シンショウ				
20					6枚
21					
22	>*				
23	//				

No. 1...JOB 文, A1234 はユーザ id, XYZ は3文字以内の任意の英数字

No. 3...ABC. DATA は格納するデータセット名 (補足1参照)

No. 4...データ種別コード (i2) (補足3参照)

No. 5...データタイトル (80文字以内)

No. 6...データ行列の行の大きさm (i5)

No. 7...列の大きさn (i5)

No. 8...データを行単位ごとに入力 (10F8.0)

No.12...行ごとのタイトルの有無 (i2) (1は有, 0は無)

No.13...行ごとのタイトル (12文字以内)

No.16...列ごとのタイトルの有無 (i2) (1は有, 0は無)

No.17...列ごとのタイトル (12文字以内)

No.20...データの終わり

// ...JOB の終わり

例2 標準外フォーマットのデータを端末から入力させる場合 (注 下線部分はユーザが入力すべき部分)

NO.	5	10	15	20	25
1	LOGON A1234/パスワード				
2	>>USE ABC.DATA				
3	>>EDIT,NEW				
4					
5	INPUT				
6					
7	00010	1.4	170.3	39.45	10.5
8	00020	0.05	155.5	32.10	12.3
9	00030	0.8	159.0	34.30	11.
10	↓				
11	02550	1.2	160.4	35.44	11.
12	02560				
13					
14	E>20	0.05	162.4	37.0	12.3
15	E>END SAVE				
16					

No. 1...TSS ジョブの開始, A1234 はユーザ id

No. 2...ABC. DATA はデータセット名

No. 3...新しくデータセットを作成することを指示

No. 7...データを行単位ごとに入力

フォーマットはユーザが自由に決めることができ、プログラム実行段階に、ダイナミックフォーマットで指定する。

No.13...空行を入力, データの終わり

No.14...2行目のデータを修正

No.15...データセットに保存

## 補足1 データセット名のつけ方

データセット名の構成は次のとおりである。

ユーザ指定名. データ識別子

ユーザ指定名は A～Z で始まる 8 文字以内の英数字または - (マイナス) からなる文字列で、ユーザが自由に指定することができる。データ識別子は必ず DATA としなければならない。

例 TAIKAKU. DATA

DA-81-4. DATA

## 補足2 データセットの消去

データセットの消去は、次のコマンドを入力する。

»FPURGE△ABC. DATA

ABC. DATA は消去するデータセット名。

## 補足3 データ種別コード

1. 観測値行列
  2. 変量間の相関行列
  3. 個体間の相関行列
  4. 変量との因子負荷行列
  5. 個体との因子負荷行列
5. 出力データファイルの作成

分析結果の出力はディスプレイ及びタイプライター上に印刷される。後に、再び入力データとして使用したいときには、この出力を直接システム側で用意したデータセットに書き込むこともできる。したがって使用者は出力したファイルについては、プログラム実行終了後、自分のデータセット名に変更しておく必要がある。

\*\* Q112 : IF YOU WANT TO STORE

INPUT-FILE TYPE 1, ELSE '0'.

## 補足1 データセット名の変更方法

データセット名の変更は、次のコマンドを入力する。

»FNAME△旧データセット名, 新データセット名

## C プログラム作成の経過

プログラム作成にあたっては、藤田倫子氏の協力を得た。作成の経過と担当は次の通りである。

## 主プログラム

プログラムの全体的構成についての原案を芝が作成し、その指示にもとづいて、藤田氏がプログラミングをおこなった。

## 副プログラム

主要な解法などの副プログラムの作成者は次の通りである。

副 プ ロ グ ラ ム	アルゴリズム	プログラミン
外れ値の検出	芝	藤田
ボックス・コックス変換	渡部, 芝	藤田
主因子法 (乗べき法)		芝
主因子法 (ヤコビ法)		芝
セントロイド法		芝
最小残差法		芝
最尤推定法	芝	芝, 藤森
A I C による因子数	芝	藤田
ジエオマックス法 (直交, 斜交)		芝
オーソマックス法		芝
一般オブリミン法		芝
プロクラステス法 (直交, 斜交)		芝
因子スコア (EQ13, EQ7)		芝
四分相関係数		塗師

## II 主なオプションの説明

プログラムは会話型となっており、端末の表示を通して要求されたデータ (オプションの選択に関する情報も含む) を、所定の方法で入力することによって、進行していく。以下、六つのコンポーネントのオプションのうちそれぞれ主なものについて、簡単に説明する。なお、説明中に出て来る式の番号は、『因子分析法(第2版)』(芝, 1979) に依る。また、式の番号に→印の付いているものは、全く同じものではないが、同等の意味をもつもの、あるいは最も近い関係にあるものを指す。

## A データの入力

\*\* Q11 : HOW TO INPUT DATA ?

1. FROM KEY BOARD
2. FROM DATA FILE (WITH STANDARD FORMAT)
3. FROM DATA FILE (WITH FREE FORMAT)

データの入力方法を指定する。

1. 今、キーボードから指示に従って直接入力する場合
2. 既にデータセットに格納済みの、標準フォーマットを持つデータを入力する場合
3. 既に格納済みの、標準外フォーマットを持つデータを入力する場合

## B プログラムの進行

```

** Q113 :   SELECT A NEXT STEP

0. GO TO COMPONENT 2 (DATA SCREENING)
1. GO TO COMPONENT 3 (COMPUTE CORRELATION)
2. SELECT DATA INPUT AGAIN
8. GO TO ANOTHER COMPONENT
9. EXIT FROM FACTOR ANALYSIS
    
```

プログラムの各段階で分岐があり、次に進む先を指定できる。ここにあげたものはその一例であるが、ときによって少しずつオプションの内容が異なる。

0. 次のコンポーネント 2 へ進む
1. 次のコンポーネント 2 を飛ばして、コンポーネント 3 へ進む
2. コンポーネント 1 でもう一度入力をおこなう
8. 他のコンポーネントへ進む
9. ジョブを終了する

なお、このオプションのうち、8 と 9 は分岐のオプションとしていちいち表示されないことがある。表示はなくとも、それを入力すれば機能するようになっている。

## C 予備的なデータ処理

因子分析の計算に入るまえに、データについて予備的な処理を行なうことができる。

```

**Q21 :   KINDS OF MANIPULATION

1. DELETION
2. CORRECTION
3. OUTLIER CHECK
4. TRANSFORMATION
0. SCREENING END
    
```

### 1. データの一部削除

データ行列から特定の変数に関する列をすべて削除するとか、特定の個体に関する行をすべて削除することができる。

```

**Q21-1 :   KINDS OF DELETION

1. DELETE VARIABLES
2. " SUBJECTS
    
```

### 2. データの修正

データ行列の名称の変更、変数名、個体名の変更、データの値の修正などができる。

```

**Q21-2 :   KINDS OF CORRECTION

1. CORRECT TITLE
2. " VARIABLE NAMES
3. " SUBJECT NAMES
4. " DATA VALUES
    
```

### 3. 外れ値の検出

データの中に、何らかの誤りによって異常に大きい（あるいは小さい）値のものが混入している恐れがあるとき、その疑いのある測定値を検出する目的で、外れ値に関する検定を行う（竹内・大橋, 1981; Barnett and Lewis, 1978; 簡約統計数値表, 1977）。一つの変数の母集団分布が正規分布をするとき、それからの標本として、最大値（あるいは最小値）がある値を越える確率が所定の値  $\alpha$  以下のとき、これを外れ値として検出する。ここでは  $\alpha = 0.01$  とし、統計量としては

$$T_n = \frac{x(n) - \bar{x}}{S}$$

あるいは、

$$T_1 = \frac{x(1) - \bar{x}}{S}$$

を用いて、検定が行われる。有意な最大値（あるいは最小値）があったとき、これを除いた  $n-1$  の標本について、あらためて外れ値の検定が行われる。

なお、外れ値の数が標本数の 10% を越えたときには、その旨を知らせ、それ以上の外れ値の検出は行なわない。このように大量の外れ値があるのは、その変数が正規分布の仮定をするのに明らかに不都合なような分布をしている可能性が考えられる。

### 4. データの変換

測定誤差が、測定値  $y$  と独立でない恐れがある場合など、データが正規分布に近似するよう変数変換をすると都合のよい場合がある。ここでは次のような変換 (Box and Tiao, 1973) だけが可能である。

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & (\lambda \neq 0) \\ \log y & (\lambda = 0) \end{cases} \quad y > 0$$

$\lambda$  の値を指定することによって、たとえば次のような変換ができる。

- $\lambda = -1$       逆数変換
- $\lambda = 0$         対数変換
- $\lambda = 1/2$       平方根変換

このほか  $\lambda$  としてどのような値を用いるのがよいかあらかじめ不明のときなど、変換後のデータの分布が正規分布によく近似するよう  $\lambda$  の最尤推定値をもとめ、これを用いて変換することもできる<sup>1)</sup>。ただし、この変換は、もとのデータが比尺度であって、意味のある原点をもつ場合でなければ適用できない。

## D 因子分析モデルの適用：R技法かQ技法か

因子分析モデルには変数間の相関行列を用いるモデルと、個体間の相関行列を用いるモデルとがある。

\*\* Q31 : HOW TO COMPUTE CORRELATION ?

1. COMPUTE BY USING R-TECHNIQUE
2. COMPUTE BY USING Q-TECHNIQUE

#### 1. R 技法による因子分析のための相関行列

相関行列として、変数間の相関係数を要素とするものを用いる通常の分析法を R 技法とよんでいる。

#### 2. Q 技法による因子分析のための相関行列

通常の因子分析法 (R 技法) のように変数間の相関係数行列をもちいるのではなく、個体間の相関係数行列を用いる分析法を、Q 技法という。(Comrey, 1973)

### E 行あるいは列ごとの標準化

\*\* Q32 : IF YOU WANT TO STANDARDIZE  
0

SUBJECT TYPE 1, ELSE '0'

R 技法の場合、ふつうは、各変数は独自の尺度によって測定されていてよく、個体ごとの標準化をする必要はないが、何らかの理由で全変数の測定値を個体ごとに標準化する必要があることがある。たとえば、すべての変数が共通する尺度 (たとえば共通する意味をもった評定尺度) によって測定されているが、個体 (被験者) ごとに尺度値の分布に大きな差異があり、しかも、これはたんに各被験者の評定値の与え方の個人差で、これが変数間のみかけ上の相関を高める恐れがある場合などである。このときには個体ごとにその平均  $\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$  とそのまわりの標準偏差  $s_i$  をもちいて、すべての測定値をつぎのように標準化しておく。

$$z_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$$

\*\* Q33 : IF YOU WANT TO STANDARDIZE

VARIABLE TYPE 1 ELSE '0'

Q 技法では個体間の相関係数をもとめるが、この場合、各変数ごとに測定の尺度が異なり不都合が生ずることが多い。そのような場合には、変数ごとに、その平均  $\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$  と、そのまわりの標準偏差  $s_j$  とを用いて、次のように標準化しておく。

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

なお、すべての変数が共通の尺度 (たとえば共通の意味をもつ 7 段階評定など) で測定されているときには、このような標準化をする必要はない。

### F 相関係数の種類

相関係数の算出には、どのような方法が用いられるか。

\*\* Q34 : KINDS OF CORRELATION

1. PRODUCT MOMENT CORRELATION MATRIX
2. TETRACHORIC CORRELATION MATRIX

#### 1. 積率相関係数

ピアソンの積率相関係数によるものである。変数が 2 値変数であたえられているときには、いわゆるファイ係数と同等である。

#### 2. 四分相関係数

二つの変数が、いずれも 2 値変数 (たとえば 0 と 1 など、二つの値だけを実現値とする変数) であって、それがともに 2 変量正規分布をする確率変数をそれぞれ特定の値で 2 分したものとみなすことができるとき、そのもとの確率変数間の相関係数の推定をしてもとめられたものが、四分相関係数である。したがって、変数が 2 値変数であらわされているが、それらが、もともと連続量で表現するのにふさわしい特性にかかわるものである場合に、その連続量を用いた因子分析を行う目的から、この四分相関係数を用いることがある。しかし、このプログラムでは一組のデータを構成するすべての変数について、一括して相関係数をもとめるようになっているので、一部の変数の間についてのみ四分相関係数を求めるということとはできない。

四分相関係数の算出には Kirk (1973) の方法が用いられているが、各変数の周辺分布に大きな偏りがある場合には、その変数の組について四分相関係数の推定は行なわれない。そのような場合には相関係数の推定ができなかった変数は削除され、より少ない数の変数について相関係数行列が構成される。

### G 因子の抽出

相関行列から初期因子解を抽出する方法のうち、以下の 5 通りの特徴的な方法が選択できる。

\*\* Q41 : METHOD OF FACTOR EXTRACTION

1. PRINCIPAL FACTOR BY POWER METHOD
2. PRINCIPAL FACTOR BY JACOBI METHOD
3. MINIMUM RESIDUAL METHOD
4. CENTROID METHOD
5. MAXIMUM LIKELIHOOD METHOD

#### 1. 主因子法 (乗ベキ法によるもの)

相関行列 (あるいは残差相関行列  $R$ ) に対する固有ベクトルとして因子負荷ベクトルを、1 因子ごとに乗ベキ法によってもとめる。次の式、

$$a_{i+1} = \frac{Ra_i}{\sqrt{a_i Ra_i}} \quad (\rightarrow 2.48)$$

によって、逐次的に近似を高めていく。因子寄与の大きな因子から順次負荷がもとめられる。したがって、はじめの数因子のみをもとめる場合には計算時間も少なく便利である。しかし、2つの因子の寄与の比が1に近いときには計算時間が増大し、大きさの等しい固有値がある場合には、この方法によって解をもとめることはできない。

## 2. 主因子法（ヤコビ法によるもの）

相関行列をヤコビ法によって次のように分解する解である。

$$R = Q\Lambda Q' \quad (\rightarrow 2.49)$$

$Q$ は規準化された固有ベクトルを各列にもつ行列、 $\Lambda$ は固有値を対角成分とする対角行列である。因子負荷は  $A = Q\Lambda^{1/2}$  によってあたえられる。ただし  $A$ は、各因子の因子負荷ベクトルを列にもつ因子負荷行列をあらわす。

この解は計算誤差を除いては、乗べき法によるものと同じ解をあたえる。ただし(2.49)のヤコビ法では等しい固有値の場合や、負の固有値の場合も解がもとめられる。因子負荷はもちろん正の固有値に対するベクトルについてのみ算出される。すべての因子について解が同時にもとめられるが、それだけ時間がかかる。

## 3. 最小残差法

順次の因子の因子負荷ベクトルをもとめていくとき、残差相関行列の非対角要素の平方和が最小となるように解をもとめる方法。共通性はあらかじめ推定せず、残差相関行列の対角成分に因子負荷ベクトルの要素の平方を入れたもの  $R^*_{p-1}$  をもちいて、

$$a_p = \frac{R^*_{p-1} a_p}{\sqrt{a_p R^*_{p-1} a_p}} \quad (\rightarrow 2.76)$$

の式によって逐次的に因子負荷  $a_p$  をもとめる。

## 4. セントロイド法

因子負荷の絶対値和  $\sum_{j=1}^n |a_{jp}|$  を最大とすることを目標とした解である。（実際には、必ずしも最大となる解がつねに与えられているとは限らない。）因子寄与が順次急激に減少していくような構造を相関行列がもっているときには、主因子法によく近似した解になる。計算時間は他の初期解に比べて著しく短い。したがって、変数の数  $n$  の大きいデータから比較的少数の因子の負荷をもとめるときなど主因子の近似解として利用される。

## 5. 最尤法

$m$  この共通因子が、 $m$  変量正規分布  $N(O, I_m)$  に従い、 $n$  この独自因子が、 $n$  変量正規分布  $N(O, I_n)$  に従

う確率変数であるという仮定のもとに、あたえられた観測値による分散共分散行列の尤度を最大とする解をもとめるものである。

最尤解では、あらかじめ指定した因子数のもとでの解があたえられているが、共通性は因子負荷と共に推定されるので、あらかじめ推定しておく必要はない。

このプログラムでは、まず独自性が推定される。その中の逐次的数値計算法ではニュートン・ラフソン法が用いられている<sup>2)</sup>。ただし、データによっては、独自性が正の値としてもとまらない変数を含むことがある。このような場合には、この変数の共通性は、1.0であると見なし、まず、この変数そのものを因子として抽出し、そのあとの残差相関行列に再び最尤法を適用する。得られた因子解において、因子負荷の値が1.0となっている変数がみられるのは、このような場合である<sup>3)</sup>。

## H 共通性の推定

主因子法（乗べき法、ヤコビ法とも）、セントロイド法の場合には、あらかじめ共通性の推定をする必要がある。

```

** 042-1 :   HOW TO ESTIMATE COMMUNALITIES

      1. SMC
      2. MAXIMUM VALUE OF NONDIAGONAL
      3. 1.0
      * 4. INPUT FROM KEY BOARD
    
```

### 1. SMC

ある変数の共通性を、その変数と他の  $n-1$  この変数との間の重相関係数の平方によって推定するもの。すなわち、共通性を成分とする対角行列  $H^2$  は

$$H^2 = I - [\text{diag } R^{-1}]^{-1} \quad (4.35)$$

によってもとめられる。

SMC による共通性の推定値は、共通性の下限になっている。因子数に比べて変数の数が大きくなるほど、この値は、よい推定値となる。

### 2. 非対角成分のうち絶対値の最大の値

便宜的によく用いられる推定値で、セントロイド法などとあわせて利用される。

3. 共通性のかわりに、相関行列の対角成分をすべて1.0とすると、いわゆる主成分分析の一種となる。

## I 因子数の推定

必要とする因子数を推定する。とくに乗べき法による主因子法、最小残差法、セントロイド法などでは、必要以上に多くの数の因子についてまで計算するのは無駄である。

**\*\* Q43 : HOW TO DECIDE THE NUMBER OF FACTORS**

1. INPUT FROM KEY BOARD
2. DEFAULT NUMBER

1. 因子数を自由に決めて、キイから入力することができる。この場合、次の 2 による因子数をもとめてみて、これを参考にして決め直してもよい。

2. 簡便法による因子数

観測数（標本数）がある程度以上（たとえば 200 以上）ある場合に、實際上意味のある因子数の目安は次の経験式によって与えられる。

$$m \cong \frac{n-2}{\log_2 n} \quad (n \geq 8)$$

≅印は右辺の値に最も近い（ただし、右辺の値より小さくない）整数をとることを示す。

**J 因子数の決定（最尤法の場合）**

最尤法の場合には、因子数を決めなければ解がもとめられない。その決め方としては、2 通りの統計的方法が提案されている。

**\*\* Q44 : HOW TO DECIDE NUMBER OF FACTORS**

1. BY STATISTICAL TEST OF SIGNIFICANCE
2. BY MINIMUM AIC ESTIMATE

1. 統計的に有意な因子数

ある因子数のもとで最尤推定された因子解により、あてはまりのよさを推定すると、その因子数が少なすぎる場合には、仮説（その因子数による因子分析モデル）が棄却される。したがって因子数を順次増して行って仮説が棄却されなくなったときの最小の因子数をもって統計的に有意な因子数とする。

この方法は、統計的に有意な残差分散・共分散を検出するものであるから、標本数が多くなるに従って有意とされる因子数が多くなる傾向がある<sup>4)</sup>。

2. AIC による因子数の決定

モデルのあてはまりの良さをあらわす指標として赤池の情報基準（AIC）をもちい、これを最小とする因子数を採用する方法である。この場合、エントロピーの比較に際して、漸近的近似がもちいられているので、標本数は十分大きいことが必要である。また、そのため、最適とされる因子数は標本数とは無関係に定まる<sup>5)</sup>。

**K 因子の変換**

因子を変換して、単純構造をもとめたり、仮説的因子構造へのあてはめを行なうことができる。

**\*\* Q51 : METHOD OF FACTOR TRANSFORMATION**

1. VARIMAX METHOD
2. ORTHOMAX METHOD
3. GEOMAX METHOD (ORTHOGONAL)
4. GEOMAX METHOD (OBLIQUE)
5. OBLIMIN METHOD
6. PROCRUSTES METHOD (ORTHOGONAL)
7. PROCRUSTES METHOD (OBLIQUE)

1. バリマックス法

初期解を直交変換して単純構造をもとめる解として最もよくもちいられるもの。因子ごとに因子負荷の平方の分散をもとめ、これらの全因子についての和を単純構造の指標とする。この指標（バリマックス基準）は

$$V = \sum_{p=1}^m \left\{ \frac{1}{n} \sum_{j=1}^n \left( b_{jp}^2 - \frac{1}{n} \sum_{k=1}^n b_{kp}^2 \right)^2 \right\} \quad (6.2)$$

によってあらわされる。これを最大とする直交変換をもとめる方法がバリマックス法である。これは、オーソマックス法において、係数  $w$  を 1.0 としたものに等しい。

2. オーソマックス法

単純構造をあたえる一連の直交解の総称。変換後の因子負荷行列を  $B = (b_{jp})$  とすると、

$$U = \sum_{p=1}^m \sum_{j=1}^n b_{jp}^4 - \frac{w}{n} \sum_{p=1}^m \left( \sum_{k=1}^n b_{kp}^2 \right)^2 \quad (6.38)$$

を最大とする解である。ここで  $w$  は定数で、 $w$  の選択によって解に若干の違いが生ずる。オーソマックス解に属する個別の解の名称と  $w$  の大きさは次のとおり。

コーティマックス	$w=0$
バイコーティマックス	$w=0.5$
バリマックス	$w=1$
エカマックス	$w=m/2$
パーシマックス	$w=n(m-1)/(m+n-2)$
因子パーシモニー	$w=n$

$w (0 < w < n)$  が小さいときには、因子寄与の偏りが大きく、一部の因子の負荷が全体的に大きくなるという傾向がある。

3. ジェオマックス法（直交）

単純構造をもとめるための直交変換解で、任意の 2 因子による平面上において、適当な方法によって 2 群にわけられた変数が、それぞれ一方の因子軸の上に、なるべく大きな射影をもつような位置に 2 因子の軸を直交回転する方法である。バリマックス法とほぼ同じ結果をあたえるが、時には因子寄与の平準化などの点でバリマックス法よりよい結果をあたえることがある。

4. ジェオマックス法（斜交）

直交のジェオマックス法と同様の原理で、単純構造をもとめる解であるが、その際に、因子軸が相互に直交するという制約を設けないため、一般に因子軸は斜交する

ことになる。

#### 5. 一般オブリン法

単純構造をあたえる一連の斜交解の総称。変換後の因子構造行列を  $S=(s_{jp})$  とすると、

$$K = \sum_{p \neq q} \left\{ \sum_{j=1}^n s_{jp}^2 s_{jq}^2 - \frac{\gamma}{n} \left( \sum_{j=1}^n s_{jp}^2 \sum_{j=1}^n s_{jq}^2 \right) \right\} \quad (7.4)$$

を最小とする解である。ここで  $\gamma$  は定数で、 $\gamma$  の選択によって、解に若干の違いが生ずる。

一般オブリンに属する個別の解の名称と  $\gamma$  の大きさは、次のとおり。

コーティミン	$\gamma=0$
バイコーティミン	$\gamma=.5$
コバリミン	$\gamma=1.0$

$\gamma(0 < \gamma < 1.0)$  の値が小さいもの程、因子は斜交化し、 $\gamma$  の値が高い程直交化する傾向がみられる。

#### 6. プロクラステス法 (直交)

何らかの情報にもとづいて、因子負荷の仮説的な構造が与えられているとき、それにもっともよく近似するように初期解の直交変換をする解。与えられた初期因子負荷行列を  $A$  として、仮説的な因子負荷行列を  $C$  とすると、プロクラステス解は、直交変換  $T$  によって、残差の平方和

$$tr(AT-C)'(AT-C)$$

を最小とする解であり、

$$B = AA'C(C'AA'C)^{-1/2} \quad (\rightarrow 9 \cdot 10)$$

によってあたえられる。

なお、仮説的因子負荷行列の要素の一部について、これを指定しないでおくこともできる。この場合には、指定しない要素の値を 2.0 以上としておけばよい。たとえば、はじめの 7 変数はそれぞれ、第 1、第 2、第 3 因子のいずれかにおいて負荷が顕著に高いことがあらかじめ予想され、第 8、第 9 変数については、第 1 因子への負荷が殆んどないことだけがわかっているとき、仮説的因子構造行列の一例として次のような行列を用いる。

$$C = \begin{pmatrix} 1.0 & 0 & 0 \\ 1.0 & 0 & 0 \\ 1.0 & 0 & 0 \\ 0 & 1.0 & 0 \\ 0 & 1.0 & 0 \\ 0 & 0 & 1.0 \\ 0 & 0 & 1.0 \\ 0 & 2.0 & 2.0 \\ 0 & 2.0 & 2.0 \end{pmatrix}$$

#### 7. プロクラステス法 (斜交)

何らかの情報にもとづいて、因子構造あるいは因子パタンの仮説的な構造が与えられているとき、それにもっ

ともよく近似するように初期解を変換する解。この場合、因子間の直交条件を設けていない。

あたえられた初期解を  $A$  とし、仮説的因子行列を  $C$  とすると因子構造に関する斜交のプロクラステス法は、両者の不一致をあらわす偏差の平方和

$$tr(S-CD)'(S-CD)$$

を最小とするものとしてあたえられる。ただし  $S$  は直交因子負荷行列  $A$  を変換した斜交因子構造  $S=AT$  をあらわし、 $D$  は仮説的因子構造の因子ごとのベクトルのノルムを調節するための比例定数を成分とする対角行列である。この因子解は

$$S=AQD \quad (\rightarrow 9 \cdot 20)$$

によって与えられる。ただし、

$$Q=(A'A)^{-1}A'C$$

$$D=\text{diag}\{Q'Q\}^{1/2} \quad (9 \cdot 22)$$

である。斜交解であるため、因子構造をあらわすこの  $S$  のほかに、因子パターン  $P$  と、因子間の相関行列  $L$  とが算出される。これらの間には、 $P=SL^{-1}$  という関係がある。

因子パターンに関する斜交プロクラステス法の場合には、因子パターンをあらわす仮説的行列  $C$  を用い、 $P=AQD$  によって解が与えられる。ただし

$$Q=(A'A)^{-1}A'C$$

$$D=\text{diag}\{Q'Q\}^{1/2}$$

である。

#### L 因子負荷の大きい順に変数を配列

単純構造をもとめた因子解の因子負荷行列を見やすくするために、変数の配列順序を入れかえて、各因子毎に因子負荷の高い変数から順次配列する。その場合、各変数は、その変数の中で最大の因子負荷を示す因子の中で順序づけられる。

```
** Q54 : IF YOU WANT DECENDING
1
```

```
ORDER LIST TYPE 1 ELSE '0'
```

#### M 仮説的行列は因子構造か因子パターンか

斜交のプロクラステス法の場合には、仮説的行列として、因子構造を与えるのか、因子パターンを与えるのかを指定する必要がある。

```
** Q58 : WHICH IS TO BE GIVEN ?
1. HYPOTHETICAL STRUCTURE
2. HYPOTHETICAL PATTERN
```

1. 仮説的行列として因子構造を与える。
2. 仮説的行列として因子パターンを与える。

### N プロクラステス法に用いる初期因子解の因子数

$m$ 次元空間における初期因子解として、 $n \times m$ の因子負荷行列が与えられているとき、これより少ない因子数( $l < m$ )の仮説構造行列への近似変換解をもとめるには、初期因子負荷行列の特定の  $l$  列を選んで用いてもよいし、これより多い因子数をもちいてもよい。このプログラムでは、初期因子解の因子数の指定がもちいる仮説構造行列の因子数より大きいときには、その因子数にあたる分だけ、仮説因子行列に新たな因子負荷の値として、2.0が加えられるようになっている。

```
** 059 : INITIAL FACTOR LOADING MATRIX CONSISTS OF 4 FACTORS,
        THEN, HOW MANY DIMENSIONS OFSPACE FACTOR DO YOU USE ?
        TYPE IT 2 DIGITS
```

### O 仮説的構造行列の入力

プロクラステス法では仮説的構造（あるいはパターン）行列を与える必要がある。

```
** 0510 : HOW TO INPUT HYPOTHETICAL STRUCTURE ?

1. FROM KEY BOARD
2. FROM DATA FILE (WITH STANDARD FORMAT)
3. FROM DATA FILE (WITH FREE FORMAT)
```

1. キーボードから入力する。
2. 標準フォーマットを持つファイルから入力する。
3. 標準外フォーマットを持つファイルから入力する。

### P 因子スコアの推定

因子解が定まったら、これに対する個体の因子スコアを算出する。因子スコアは、はじめに与えられたデータ行列を用いて、各変数の観測値の一次結合によって推定される。

なお、当然のことであるが、はじめのデータ行列が与えられていない場合、あるいは、与えられていても四分相関係数を用いた場合などには因子スコアをもとめることはできない。

```
** 061 : HOW TO ESTIMATE FACTORS ?

1. BY EQUATION NO.7
2. BY EQUATION NO.13
```

#### 1. 因子スコアの推定 $F_7$

共通因子によってあらわされる変動分  $FP'$  について、

推定値  $\hat{F}P'$  がもっともよく一致するよう、因子スコアの推定を行なう。誤差

$$E = FP - \hat{F}P \quad (\rightarrow 10 \cdot 21)$$

の平方和を最小とする解である。因子スコアを推定するための重み係数の行列  $W$  は、

$$W = R^{-1}(R - D^2)\hat{P}(\hat{P}'\hat{P})^{-1} \quad (\rightarrow 10 \cdot 23)$$

によって、そして因子スコアの推定値は、 $F_7 = ZW$  によってあたえられる。ここで  $\hat{P}$  は解としてあたえられた因子パターンをあらわす。直交解の場合には因子パターンは因子負荷にひとしい。

#### 2. 因子スコア $F_{13}$

因子解は因子負荷（あるいは因子パターン）によってあたえられるのがふつうである。これに対し、共通因子スコア  $F$  は、独自性因子スコア  $U$  が未知であるため、確定的にもとめることはできない。何らかの基準にもとづいて誤差を最小とするような因子スコアの推定値  $\hat{F}$  がもとめられる。

$F_{13}$  は、推定された因子スコア  $\hat{F}$  が真の因子スコア  $F$  にもっともよく合致するように回帰推定するもので、誤差

$$E = F - \hat{F} \quad (10 \cdot 10)$$

の平方和を最小とする解である。因子スコアを推定するための重み係数の行列  $W$  は

$$W = R^{-1}S \quad (10 \cdot 14)$$

によってあたえられ、因子スコアは  $F_{13} = ZR^{-1}S$  によって推定される。ここで  $S$  は解としてあたえられた因子構造（直交解の場合には因子負荷）をあらわす。

#### 注 1 ボックス・コックス変換

観測値  $y(>0)$  を

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} & (\lambda \neq 0) \\ \log y & (\lambda = 0) \end{cases}$$

によって変換したとき、 $y^{(\lambda)}$  の分布が正規変数  $N(\mu, \sigma^2)$  の分布に最もよく近似するよう  $\lambda$  を定める。ここで  $\mu, \sigma^2$  は観測値に依存して決まる定数である。変換された  $n$  の値  $y_i^{(\lambda)}$  の尤度は、

$$L = \frac{J}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i^{(\lambda)} - \mu)^2\right\}$$

とあらわされる。ここで  $J$  は変換のヤコビアン

$$J = \prod_{i=1}^n \frac{dy_i^{(\lambda)}}{dy_i} = \prod_{i=1}^n y_i^{\lambda-1} = \hat{y}^{n(\lambda-1)}$$

である。ただし、 $\hat{y}$  は観測値の幾何平均  $\hat{y} = \left(\prod_{i=1}^n y_i\right)^{1/n}$

をあらわす。まず、この尤度を最大とする  $\mu$  をもとめるために、式の対象を  $\mu$  に関して微分し、その結果を 0 とする  $\mu$  の値をもとめると、

$$\mu = \frac{1}{n} \sum_{i=1}^n y_i^{(2)} = \bar{y}^{(2)}$$

となる。また、 $\sigma^2$  に関して  $L$  を微分し、その結果を 0 とする  $\sigma^2$  をもとめると、

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (y_i^{(2)} - \bar{y}^{(2)})^2$$

となる。これらを、尤度をあらわす上式の右辺に代入すると、

$$L = (2\pi)^{-2/n} J \cdot \left\{ \frac{1}{n} \sum_{i=1}^n (y_i^{(2)} - \bar{y}^{(2)})^2 \right\}^{-n/2} \exp \left\{ -\frac{n}{2} \right\}$$

となる。この値を最大とする  $\lambda$  は

$$\sum_{i=1}^n \left\{ \frac{y_i^{(2)} - \bar{y}^{(2)}}{J^{1/n}} \right\}^2$$

を最小とする  $\lambda$  の値としてあたえられる。したがって  $z_i^{(2)} = y_i^{(2)} / J^{1/n}$  とすると、最適な  $\lambda$  は、

$$\sum_{i=1}^n (z_i^{(2)} - \bar{z}^{(2)})^2$$

を最小とする  $\lambda$  としてあたえられることになる。ここで  $\bar{z}^{(2)}$  は  $z_i^{(2)}$  の平均である。

実際に  $\lambda$  の値を計算するためには次のような変数  $w_i = y_i / \bar{y}$  をもちいるとよい。これによると

$$z_i^{(2)} = \begin{cases} \bar{y} \lambda^{-1} w_i^{\lambda} + C_{\lambda} & (\lambda \neq 0) \\ \bar{y} \log w_i + C_0 & (\lambda = 0) \end{cases}$$

となる。ただし  $C_{\lambda}$ ,  $C_0$  は  $\lambda$  と  $\bar{y}$  の関数

$$C_{\lambda} = -\frac{1}{\lambda \bar{y}^{\lambda-1}}$$

$$C_0 = \bar{y} \log \bar{y}$$

で、いずれも  $w_i$  を含まない。したがって、第 1 項の値  $\bar{y} \lambda^{-1} w_i^{\lambda}$  (あるいは  $\lambda=0$  のときは  $\bar{y} \log w_i$ ) について、その平均からの偏差の平方和を最小とする  $\lambda$  を数値的方法によって探してゆけばよい。

## 注 2 最尤因子解

因子数  $m$  に対する因子解  $\hat{A}$  は、独自性  $\Psi$  があたえられれば

$$\hat{A}_m = \Psi Q_m (A_m - I)^{1/2} \quad (\rightarrow 11.88)$$

によってもとめることができる。ただし、 $A_m$  は、 $\Psi^{-1/2} S \Psi^{-1/2}$  のはじめの  $m$  個の固有値を成分とする対角行列を、 $Q_m$  はそれに対応する規準化された固有ベクトルを列にもつ行列をあらわす。このような条件のもとで、尤度  $L$  を最大とする独自性  $\Psi$  の推定値をもとめるとというのが、ここでもちいられた解法の特徴である。

その際に以下のように、ニュートン・ラフソン法がも

ちいられる。いま  $\Psi$  のまわりでつぎのように微分  $L'(\Psi)$  のテイラー展開をする。

$$L'(\Psi + \Delta\Psi) = L'(\Psi) + L''(\Psi) \Delta\Psi + H$$

ここで、 $L'$  は関数  $L$  の、 $\Psi$  の各対角成分に関する微分  $\frac{\partial L}{\partial \phi_i}$  を成分とする大きさ  $n$  の対角行列をあらわし、 $L''(\Psi)$  は  $\Psi$  の各対角成分による 2 次微分をあらわし、次のような大きさ  $n \times n$  の、いわゆるヘシアン行列となる。

$$L''(\Psi) = \left( \frac{\partial^2 L}{\partial \phi_i \partial \phi_k} \right)$$

また、右端の  $H$  は、2 次以上の高次の項をまとめてあらわしたものである。いま  $H$  を無視できるものとする  $\Psi$  のある近似値に対して  $L'(\Psi + \Delta\Psi) = 0$  となるような  $\Delta\Psi$  は、 $L''(\Psi)$  が正則であるとして、

$$\Delta\Psi = -L'(\Psi) [L''(\Psi)]^{-1}$$

によってあたえられる。従って  $\Psi$  の最尤推定値は  $\Psi_0 = \Psi - L'(\Psi) [L''(\Psi)]^{-1}$  によって近似される。この近似が十分良いためには、無視された高次の項  $H$  の影響が十分に小さくなければならない。実際にはこの近似式をくりかえし適用することによって、推定値の近似を高めていくという方法がとられる。

さて、実際にこのような方法を実行するには、 $L'(\Psi)$  および  $L''(\Psi)$  をもとめなければならない。 $L'(\Psi)$  は次のようにして求められる。

$$\frac{\partial L}{\partial \Psi} = \text{diag}(\Sigma^{-1} - \Sigma^{-1} S \Sigma^{-1}) \quad (\rightarrow 11.91)$$

$L''(\Psi)$  は、これから直接もとめるのではなく、つぎのようにしてもとめる。まず、因子解  $\hat{A}$  による母分散共分散の推定値を  $\hat{\Sigma} = \hat{A} \hat{A}' + \Psi$  すると、 $m$  因子の解の場合の尤度の 1 次微分は、

$$L'_m(\Psi) = \text{diag} \{ \Psi^{-1} (\hat{\Sigma} - S) \Psi^{-1} \} \quad (\rightarrow 11.94)$$

となる。ここでさらに、さきの因子解  $\hat{A}$  に関する条件 (11.88) を代入して整理すると、 $L'_m(\Psi)$  の対角成分は、

$$\frac{\partial L_m}{\partial \phi_i} = \frac{1}{\phi_i} \left\{ \sum_{k=m+1}^n (1 - \lambda_k) q_{ik}^2 \right\}$$

となる。これを  $\Psi_j$  で微分することによってヘシアン行列が得られる。それは、

$$L_m''(\Psi) = \Phi^{(2)} - 2L_m'(\Psi) \Psi^{-1} + V$$

となる。ここで  $\Phi^{(2)}$  は次の行列  $\Phi$  の各成分の平方を成分とする行列をあらわす。

$$\Phi = \Psi^{-1/2} (I - Q_m Q_m') \Psi^{-1/2}$$

また、行列  $V$  は

$$V = \left( \frac{1}{\phi_i \phi_j} \sum_{k=1}^m q_{ik} q_{jk} \sum_{l=m+1}^n \frac{2\lambda_k (\lambda_l - 1)}{\lambda_l - \lambda_k} q_{il} q_{jl} \right)$$

によってあたえられる。(Clarke, 1970)

以上によって  $L'_m(\Psi)$ ,  $L''_m(\Psi)$  が数値的にもとめられるので、これらをもちいて、逐次的に  $\Psi$  の最尤値をもとめていけばよい。その過程で固有値  $A_m$  や固有ベクトル  $Q_m$  は、 $\Psi$  の近似値が更新されるたびに新たに  $\Psi^{-1/2} S \Psi^{-1/2}$  を計算しなおしてもとめなければならない。因子解  $A$  の推定値は、最終的な  $\Psi$  の推定値が定まってから、それをもちいて計算する。なお、 $\Psi$  の初期値としては、

$$\phi_i = \frac{1-m/2n}{s^{(ii)}}$$

をもちいている。(Jöreskog, 1967) ここで  $s^{(ii)}$  は標本分散共分散行列の逆行列  $S^{-1}$  の対角成分をあらわす。また、計算途中において  $\Psi$  の対角成分の中に 0.001 より小さいものが出たときには、これをすべて 0.001 でおきかえる。くり返し計算中、同じ成分についてこのようなことが 5 回まで起こったときには、この変数の共通性を 1.0 として、まず、この変数を第 1 因子として抽出するものとし、そのあとの残差行列にあらためて最尤法を適用するものとする。(注 3 参照)

### 注 3 不適解の処置

最尤因子解の計算をすすめるときに、独自性の推定値が負(あるいは 0)となることがある。これは因子分析の解として不適であり、そのままでは以後の計算が不能となる。実際の逐次計算の過程で、ある変数の独自性が  $\phi_i \leq 0$  となったら、この独自性を仮の値 0.001 でおきかえ、計算をすすめる。そして以後ふたたび  $\phi_i \leq 0$  となることがなければ、それで解がもとめられる。しかし、そのようなおきかえにもかかわらず、 $\phi_i \leq 0$  がひきつづき(5 回以上)生じたら、その変数の独自性の値を 0 とみなし、(すなわち、共通性を 1.0 とし)、その変数を 1 つの因子とみなす。それにともない、この因子への他の変数の負荷を算出する。それは、この変数  $i$  と他の変数との相関係数にはかならない。こうして、第 1 因子負荷を定めたら、残差相関行列  $R_1 = R - r_i r_i'$  をもとめる。この  $R_1$  の第 1 行と第  $i$  列はすべて 0 となる。そこで、この行と列とを除いた  $n-1$  変数間の残差相関行列をもちいて、あらためて  $m-1$  因子の最尤解をもとめる。 $n-1$  変数について  $m-1$  因子の解がえられたら、さきを除いた変数  $i$  の因子負荷をあわせ、全体の大きさ  $n \times m$  の因子負荷行列をつくる。このとき、変数  $i$  の第 2 以下の因子負荷はすべて 0 となる。

### 注 4 因子数に関する統計的仮説検定

$m$  個の共通因子が存在するという仮説が真であるとき

に、標本統計量

$$\chi^2 = N' \{ \ln |\hat{\Sigma}| / |S| + \text{tr}(\hat{\Sigma}^{-1} S) - n \}$$

は、標本数が大きければ、自由度

$$\text{d.f.} = \frac{1}{2} \{ (n-m)^2 - n + m \} \quad (\rightarrow 11 \cdot 105)$$

のカイ二乗分布に近似した分布に従う。ここで

$$N' = N - 1 - \frac{1}{6}(2n+5) - \frac{2}{3}m$$

である。これをもちいて、因子数の仮説検定をおこなう。棄却されない因子数のうち、最小のものをもって因子数の推定値とする。この場合、ある因子数が真であるにもかかわらず、データを説明する因子数としては不十分であるとして、その仮説を棄て、それより大きい因子数を採用するという誤り(第一種の誤り)に対しては寛容にのぞみ、仮説の因子数よりも、真の因子数が多いにもかかわらず、仮説を採用してしまう誤り(第二種の誤り)をできるだけ避けることにし、また、簡便さをも考慮して、棄却限界値をカイ二乗分布の期待値にとるようになっている。

### 注 5 AIC による因子数の決定

いま、ある変数  $x$  が分布  $g(x)$  に従うものとする。 $x$  は一般に確率変数ベクトルをあらわす。これをパラメタ  $\theta$  をもつモデルによってあらわしたときの分布を  $f(x|\theta)$  とする。このとき、モデルのあてはまりのよさをあらわす一つの指標として、次のような確率分布のエントロピーを用いることができる。

$$B = - \int g(x) \log \frac{g(x)}{f(x|\theta)} dx$$

これは、モデルによる分布  $f(x|\theta)$  が真の分布  $g(x)$  に一致するとき 0 となり、一致しないときには負の値となる。実際には、真の分布は知られないので、このエントロピーを直接求めることはできないが、2 つのモデルについて、あてはまりの良さを比較するときには、

$$B_1 - B_2 = \int g(x) \log \frac{f_1(x|\theta)}{f_2(x|\theta)} dx$$

の正負を調べればよい。標本数  $N$  が十分に大きいときには、適当な条件のもとで最尤推定値のパラメタ  $\hat{\theta}$  によるモデルの分布のエントロピーが

$$\int g(x) \log f(x|\hat{\theta}) = \frac{1}{N} l(\hat{\theta}) - \frac{p}{N}$$

によって近似的にあらわされる(竹内, 1976)。ここで、 $l(\hat{\theta})$  は最大対数尤度をあらわし、 $p$  はモデルにおいて推定された自由なパラメタの数をあらわす。これを利用すると、2 つのパラメタのあてはまりの良さの比較は

$$B_1 - B_2 = \frac{1}{N} \{ l_1(\hat{\theta}) - l_2(\hat{\theta}) - (p_1 - p_2) \}$$

によってあらわされる。因子分析の場合には最大対数尤度は

$$l(\hat{A}, \hat{\Psi}) = -\frac{N}{2} \ln |\hat{\Sigma}| + C$$

となる。ここで  $\hat{\Sigma}$  は推定された因子解  $\hat{A}$  と独自性  $\hat{\Psi}$  とから再生される分散共分散行列をあらわし、 $C$  はモデルの比較に無関係な項をまとめたものである。

$m_1, m_2$  をそれぞれ 2 つのモデルにおける因子数とすると、自由パラメタの数は

$$p_1 = n(m_1 + 1) - m_1(m_1 - 1)$$

$$p_2 = n(m_2 + 1) - m_2(m_2 - 1)$$

となる。これらにより、2 つの因子数によるモデルのあてはまりの良さは、

$$B_1 - B_2 = \frac{1}{2} \ln \{ |\hat{\Sigma}_2| / |\hat{\Sigma}_1| \} - \frac{1}{N} (p_1 - p_2)$$

となる。

#### <文 献>

- 赤池弘次 1976 情報量規準とは何か. 数理科学 No. 153, 5-11
- Barnett, V. and Lewis, T. 1978 Outliers in Statistical Data. John Wiley
- Box, G.E.P. and Tiao, G.C. 1973 Bayesian inference in statistical analysis. Addison-Wesley
- Clarke, M.R.B. 1970 A rapidly convergent method for maximum likelihood factor analysis. British J. of Mathematical and Statistical Psychology, 23. part 1, 43-52
- Comrey, A.L. 1973 A first Course in Factor Analysis. Academic Press (芝祐順訳 因子分析法入門・サイエンス社)
- Jöreskog, K.G. 1967 Some contributions to maximum Likelihood factor analysis. Psychometrika 32, 443-482
- 簡約統計数値表, 1977, 日本規格協会
- Kirk, D.B. 1973 On the numerical approximation of bivariate normal (tetrachoric) correlation coefficient. Psychometrika 38, 259-268
- 芝 祐順 1979 因子分析法 (第2版), 東大出版会
- 竹内 啓 1976 情報統計量の分布とモデルの適切さの規準, 数理科学, No. 153, 12-18
- 竹内 啓・大橋靖雄 1981 統計的推測—2 標本問題, 日本評論社

#### 謝 辞

このプログラムの作成にあたっては、藤田倫子さんの御協力を得ました。藤田さんの秀れたプログラミングの力と、熱意によって、このプログラムが出来上ったことを記し、感謝の意を表わしたいと思います。また、このほか、一部のアルゴリズムの検討やプログラミングにおいて、塗師斌氏 (大学入試センター)、渡部洋氏 (大学入試センター)、藤森進氏 (本学大学院生) の協力を得たことを感謝します。

本研究は昭和55年度、および56年度の文部省科学研究費 (試験研究581007) の補助を受け、計算には、東京大学大型計算センターを利用しました。