



# 因子分析と共分散構造分析のパラメータ推定に関する記述統計的アプローチ

丸山, 祐造

---

**(Citation)**

国民経済雑誌, 227(4):121-133

**(Issue Date)**

2023-06-10

**(Resource Type)**

departmental bulletin paper

**(Version)**

Version of Record

**(JaLCD0I)**

<https://doi.org/10.24546/0100482445>

**(URL)**

<https://hdl.handle.net/20.500.14094/0100482445>



# 国民経済雑誌

THE  
KOKUMIN-KEIZAI ZASSHI  
(JOURNAL OF ECONOMICS & BUSINESS ADMINISTRATION)

因子分析と共分散構造分析の  
パラメータ推定に  
関する記述統計的アプローチ

丸 山 祐 造

国民経済雑誌 第227巻 第4号 抜刷

2023年6月

神戸大学経済経営学会

# 因子分析と共分散構造分析のパラメータ推定に関する記述統計的アプローチ

丸 山 祐 造<sup>a</sup>

因子分析・共分散構造分析は、主に心理統計学において発展してきた多変量データの分析手法であり、これらの手法は経営学、マーケティングを含む社会科学分野に広く浸透している。統計学では、データを所与の数値と見なしてその特徴を記述することを記述統計といい、観測データの背後に確率モデルを想定して、データを確率変数の実現値と見なす立場を推測統計という。前者の見方で基礎を固めてから、後者に進むのが通常である。しかし、因子分析・共分散構造分析は、統計学関連の講義において回帰分析よりも後で扱われるケースが多いために、記述統計的な説明はスキップされ、最初から推測統計的に扱われることが多い。本論文では、そのギャップを埋めるべく因子分析と共分散構造分析のパラメータ推定に関する記述統計的アプローチを試みる。

キーワード 因子分析，共分散構造分析，記述統計

## 1 はじめに

因子分析・共分散構造分析は、主に心理統計学において発展してきた多変量データの分析手法である。因子分析は、観測される複数の変数が、共通因子と呼ばれる少数の観測されない潜在的な因子によって記述されるというモデルに基づく分析手法である。共分散構造分析では、共通因子間の回帰関係など、因子分析より複雑なモデルを記述することが出来る。これらの分析手法は経営学、マーケティングを含む社会科学分野に広く浸透している。

統計学では、データを所与の数値と見なしてその特徴を記述することを記述統計といい、観測データの背後に確率モデルを想定して、データを確率変数の実現値と見なす立場を推測統計という。前者の見方で基礎を固めてから、後者に進むのが通常である。推測統計においては、確率モデルの想定により、そのモデルが正しい場合あるいはその近傍にある場合の精緻な理論展開が可能になる。

a 神戸大学大学院経営学研究科，maruyama@port.kobe-u.ac.jp

2次元データ  $(x_1, y_1), \dots, (x_n, y_n)$  に対する単回帰分析に対して、記述統計と推測統計のデータの理解の仕方は以下ようになる。 $Q(\alpha, \beta) = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$  を最小にする最小二乗解が

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{\alpha} = \bar{y} - \bar{x}\hat{\beta}$$

であること、 $\hat{\beta}$  は2次元データ  $(x_1, y_1), \dots, (x_n, y_n)$  を  $(x, y)$  平面に打点した散布図において、最も良く当てはまる直線を引くときの傾きである、という直感的・視覚的な理解が記述統計である。一方推測統計においては

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2) \quad \text{for } i=1, \dots, n$$

という確率モデルにおいて、確率変数  $Y_i$  の実現値が  $y_i$  であると想定する。この立場では、 $\beta$  の最小二乗推定量は確率変数であり、

$$\hat{\beta} \sim N(\beta, \sigma^2 / \sum_{i=1}^n (x_i - \bar{x})^2) \quad (1)$$

に従う。回帰係数の有意性検定は(1)のような確率分布に基づいて構成される。

因子分析・共分散構造分析は、統計学関連の講義において回帰分析よりも後で扱われるケースが多いために、記述統計的な説明はスキップされ、最初から推測統計的に扱われることがほとんどである。例えば、市川(2010)は因子分析を推測統計的に扱った良書であるが、統計学を得意としない学生にとって理解が困難であろう。本論文では、そのギャップを埋めるべく因子分析と共分散構造分析のパラメータ推定に関する記述統計的アプローチを試みる。心理統計学分野で定評ある教科書である南風原(2002)の第10章「因子分析と共分散構造分析」に沿い、その副読本として利用可能となっている。

因子分析の対象として、 $n$ 人の大学生の $p$ 個の性格特性

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} = (\mathbf{y}_1 \mathbf{y}_2 \cdots \mathbf{y}_p)$$

のようなデータを想定する。南風原(2002)は、温和、陽気、外向的、親切、社交的、協力的、積極的、素直の8つの性格特性の評定値に基づいて説明されている。第 $j$ 性格特性に関する標準化ベクトルは、算術平均を  $\bar{y}_j = \sum_i y_{ij} / n$  とするとき、

$$\tilde{\mathbf{y}}_j = \frac{1}{\sqrt{\sum_i (y_{ij} - \bar{y}_j)^2 / n}} \begin{pmatrix} y_{1j} - \bar{y}_j \\ y_{2j} - \bar{y}_j \\ \vdots \\ y_{nj} - \bar{y}_j \end{pmatrix} \quad (2)$$

で与えられる。(2)のような $n$ 次元標準化ベクトルは

$$\mathbf{1}_n^T \tilde{\mathbf{y}}_j = 0, \frac{\|\tilde{\mathbf{y}}_j\|^2}{n} = 1 \text{ for } j=1, \dots, p$$

を満たす。ただし  $\mathbf{1}_n$  は全ての成分が1の  $n$  次元ベクトル、 $\|\cdot\|$  はユークリッドノルムである。第  $j$  性格特性と第  $k$  性格特性の標本相関係数は、 $\tilde{\mathbf{y}}_j$  と  $\tilde{\mathbf{y}}_k$  の内積を用いて

$$r_{jk} = \frac{\sum_i (y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k)}{\sqrt{\sum_i (y_{ij} - \bar{y}_j)^2 \sum_i (y_{ik} - \bar{y}_k)^2}} = \frac{1}{n} \tilde{\mathbf{y}}_j^T \tilde{\mathbf{y}}_k$$

表1 性格特性の相関係数行列

温和	1							
陽気	0.03	1						
外交的	0.32	0.64	1					
親切	0.46	0.25	0.33	1				
社交的	0.27	0.53	0.88	0.36	1			
協力的	0.61	0.2	0.24	0.43	0.25	1		
積極的	0.23	0.52	0.75	0.4	0.74	0.34	1	
素直	0.42	0.42	0.33	0.45	0.27	0.46	0.24	1
	温和	陽気	外交的	親切	社交的	協力的	積極的	素直

で与えられる。従って  $p$  個の性格特性に関する  $p \times p$  標本相関係数行列は、

$$R = \frac{1}{n} \begin{pmatrix} \tilde{\mathbf{y}}_1^T \\ \tilde{\mathbf{y}}_2^T \\ \vdots \\ \tilde{\mathbf{y}}_p^T \end{pmatrix} (\tilde{\mathbf{y}}_1 \tilde{\mathbf{y}}_2 \cdots \tilde{\mathbf{y}}_p) = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix} \quad (3)$$

と書ける。表1は南風原 (2002) の表10.1の相関係数行列である。

因子分析では、第  $j$  性格特性の標準化ベクトル  $\tilde{\mathbf{y}}_j$  ( $j=1, \dots, p$ ) が、少数の (ここでは  $m$  個) 共通特性 (共通因子)  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$  と、独自特性 (独自因子)  $\mathbf{e}_j$  ( $j=1, \dots, p$ ) の線形和

$$\mathbf{z}_j = b_{j1}\mathbf{u}_1 + \cdots + b_{jm}\mathbf{u}_m + a_j\mathbf{e}_j \quad (4)$$

を用いて近似

$$\tilde{\mathbf{y}}_j \approx \mathbf{z}_j$$

出来ると考える。共通因子の係数  $b_{j\ell}$  ( $j=1, \dots, p, \ell=1, \dots, m$ ) は因子負荷とよばれる。 $\mathbf{u}_\ell$  ( $\ell=1, \dots, m$ ) や  $\mathbf{e}_j$  ( $j=1, \dots, p$ ) は全て観測されないため、それらの線形和である  $\mathbf{z}_j$  ( $j=1, \dots,$

$p$ ) も観測されない潜在変数であることに注意する。

因子分析には、大きく分けて探索的アプローチ (exploratory approach) と確認的アプローチ (confirmatory approach) の2つがある。探索的アプローチは、複数の観測変数間の相関関係が、「いくつの」「どのような内容の」因子を導入すれば説明できるのかを、探索的に調べることを目的とする。一方確認的アプローチは、因子数  $m$  および因子負荷  $b_{j\ell}$  について仮説的なモデル (一部の  $b_{j\ell}$  を 0 に固定するなど) を用意し、そのモデルをデータによって検証することを目的とする。

本論文の構成は以下の通りである。2節で探索的アプローチを、3節で確認的アプローチを扱う。特に因子負荷の推定における最小二乗法の説明に重点を置く。4節では、確認的アプローチと関係の深い共分散構造モデルについて説明する。

## 2 探索的アプローチ

### 2.1 無相関な共通因子、独自因子による $\hat{\mathbf{y}}_j$ の近似

本節では、共通因子  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$  は、全て標準化されていて互いに無相関を仮定する。つまり

$$\begin{aligned} \mathbf{1}_n^\top \mathbf{u}_\ell &= 0, \frac{\|\mathbf{u}_\ell\|^2}{n} = 1 \text{ for } \ell = 1, \dots, m \\ \mathbf{u}_\ell^\top \mathbf{u}_q &= 0 \text{ for } \ell \neq q \end{aligned} \quad (5)$$

である。(5)より共通因子間の  $m \times m$  相関係数行列は、

$$\frac{\mathbf{U}^\top \mathbf{U}}{n} = \mathbf{I}_m \quad (6)$$

となる。ただし、 $\mathbf{U}$  は  $n \times m$  行列 ( $\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_m$ ) であり、 $\mathbf{I}_m$  は  $m$  次の単位行列である。独自因子  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$  は全て標準化されていて、互いに無相関を仮定する。つまり

$$\begin{aligned} \mathbf{1}_n^\top \mathbf{e}_j &= 0, \frac{\|\mathbf{e}_j\|^2}{n} = 1 \text{ for } j = 1, \dots, p \\ \mathbf{e}_j^\top \mathbf{e}_k &= 0, \text{ for } j \neq k \end{aligned} \quad (7)$$

である。さらに  $\mathbf{e}_j$  と  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$  との無相関も仮定する。つまり

$$\mathbf{e}_j^\top \mathbf{u}_\ell = 0, \text{ for } j = 1, \dots, p, \ell = 1, \dots, m \quad (8)$$

である。(4)で与えられる共通因子と独自因子の線形和  $\mathbf{b}_j$  は、因子負荷のベクトル

$$\mathbf{b}_j = (b_{j1} \ b_{j2} \ \dots \ b_{jm})^\top \in \mathbb{R}^m$$

と  $\mathbf{U} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_m)$  を用いて、

$$\mathbf{z}_j = \mathbf{U} \mathbf{b}_j + a_j \mathbf{e}_j \quad (9)$$

と書ける。因子負荷による  $p \times m$  行列  $(\mathbf{b}_1 \ \mathbf{b}_2 \ \dots \ \mathbf{b}_p)^\top$  は因子負荷行列とよばれる。

$\mathbf{z}_j$  は標準化ベクトルの線形和なので、 $\mathbf{1}_n^\top \mathbf{z}_j = 0$  を満たす。また、(6), (7), (8), (9)よ

り

$$\frac{\|\mathbf{z}_j\|^2}{n} = \frac{\|\mathbf{U}\mathbf{b}_j + a_j\mathbf{e}_j\|^2}{n} = \|\mathbf{b}_j\|^2 + a_j^2 \quad (10)$$

なので、以下では

$$\|\mathbf{b}_j\|^2 + a_j^2 = 1 \quad (11)$$

という制約を課して、 $\mathbf{z}_j$  を標準化ベクトルとして扱う。また  $\mathbf{z}_j$  と  $\mathbf{z}_k$  の相関係数 ( $j \neq k$ ) は、(10)と同様に

$$\frac{\mathbf{z}_j^T \mathbf{z}_k}{n} = \frac{1}{n} (\mathbf{b}_j^T \mathbf{U}^T + a_j \mathbf{e}_j^T) (\mathbf{U}\mathbf{b}_k + a_k \mathbf{e}_k) = \mathbf{b}_j^T \mathbf{b}_k$$

となるので、 $\mathbf{z}_1, \dots, \mathbf{z}_p$  の相関係数行列は、

$$\frac{1}{n} \begin{pmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \vdots \\ \mathbf{z}_p^T \end{pmatrix} (\mathbf{z}_1 \ \mathbf{z}_2 \ \dots \ \mathbf{z}_p) = \begin{pmatrix} 1 & \mathbf{b}_1^T \mathbf{b}_2 & \dots & \mathbf{b}_1^T \mathbf{b}_p \\ \mathbf{b}_2^T \mathbf{b}_1 & 1 & \dots & \mathbf{b}_2^T \mathbf{b}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_p^T \mathbf{b}_1 & \mathbf{b}_p^T \mathbf{b}_2 & \dots & 1 \end{pmatrix} \quad (12)$$

で与えられる。

特に(12)の  $(j, j)$  対角成分の関係

$$\frac{\|\mathbf{z}_j\|^2}{n} = \|\mathbf{b}_j\|^2 + a_j^2 = 1 \quad (13)$$

において、 $\|\mathbf{b}_j\|^2$  は  $\mathbf{z}_j$  のうち共通因子で説明できる割合であり、 $\mathbf{z}_j$  の共通性とよばれる。一方、 $a_j^2$  は共通因子で説明できない割合であり、 $\mathbf{z}_j$  の独自性とよばれる。 $\mathbf{z}_j$  と共通因子  $\mathbf{u}_\ell$  の相関係数は、(4)、(5)、(8)より

$$\frac{\mathbf{z}_j^T \mathbf{u}_\ell}{n} = \sum_{q=1}^m b_{jq} \frac{\mathbf{u}_q^T \mathbf{u}_\ell}{n} + a_j \frac{\mathbf{e}_j^T \mathbf{u}_\ell}{n} = b_{j\ell} \quad (14)$$

であり因子構造とよばれる。因子構造による  $p \times m$  行列  $(\mathbf{b}_1 \ \mathbf{b}_2 \ \dots \ \mathbf{b}_p)^T$  は因子構造行列とよばれる。共通因子の直交性のもとで、因子構造(行列)は因子負荷(行列)に一致するが、共通因子間に相関がある場合には異なる(2.3節参照)。また因子構造行列の第  $\ell$  列の  $p$  個の要素の二乗和

$$\sum_{j=1}^p b_{j\ell}^2 \quad (15)$$

は共通因子  $\mathbf{u}_\ell$  の寄与とよばれる。寄与が大きい因子  $\mathbf{u}_\ell$  は、潜在変数たち  $\mathbf{z}_1, \dots, \mathbf{z}_p$  と高い相関を持つことを意味する。複数の共通因子が想定されている場合 ( $m \geq 2$ )、寄与に関して左側から降順に共通因子が並び替えられて、因子構造行列に対応する表が作成される。

ここで、潜在変数の線形和(4)の係数である因子負荷の推定問題を考える。因子分析の最小二乗法は、近似  $\hat{\mathbf{y}}_j \approx \mathbf{z}_j$  ( $j=1, \dots, p$ ) の精度を、2つの相関係数行列

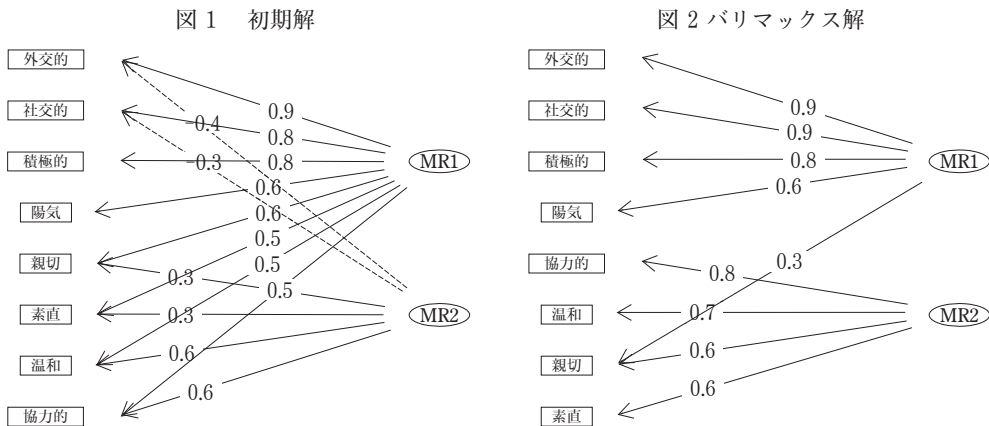
- データ  $\hat{\mathbf{y}}_j$  ( $j=1, \dots, p$ ) から計算される相関係数行列(3)

● 潜在変数  $z_j$  ( $j=1, \dots, p$ ) から計算される相関係数行列(12)の差のフロベニウスノルムの二乗

$$\sum_{j \neq k} (\mathbf{b}_j^T \mathbf{b}_k - r_{jk})^2 = 2 \sum_{j < k} (\mathbf{b}_j^T \mathbf{b}_k - r_{jk})^2 \tag{16}$$

を最小にする因子負荷  $\hat{b}_j$  ( $j=1, \dots, p$ ) を求める方法である。ただし、(11)より  $\|\mathbf{b}_j\| \leq 1$  for  $j=1, \dots, p$  の制約のもとで最小化が行われる。また(11)より

$$a_j = \sqrt{1 - \|\hat{\mathbf{b}}_j\|^2}$$



で  $a_j$  for  $j=1, \dots, p$  が推定される。表 1 の相関係数行列により推定された因子負荷とともに、共通因子と観測変数を可視化したのが図 1 である。

(13)で定義された共通性・独自性、(14)で定義された因子構造、(15)で定義された寄与は、全て因子負荷の関数である。推定された因子負荷  $\hat{b}_{je}$  を代入することで、共通性、独自性、因子構造、寄与が推定出来る。

### 2.2 直交解、解の不定性と単純構造

(5)のように共通因子  $u_1, \dots, u_m$  の無相関性 (ベクトルの直交性) のもとで得られる解は直交解とよばれる。前節で述べた解  $\hat{b}_j$  は直交解である。本節では、直交解の不定性を説明する。

任意の  $m \times m$  直交行列  $G$  ( $G^T G = I_m$  を満たす) について、線形和(9)の  $U b_j$  について

$$U b_j = U I_m b_j = U G^T G b_j = \{U G^T\} \{G b_j\}$$

と書く。このとき  $m \times m$  行列  $U G^T$  を  $m$  次元縦ベクトルの  $m$  個の組

$$U G^T = (u_{*1} \ u_{*2} \ \dots \ u_{*m})$$

と見ると、 $u_{*1}, u_{*2}, \dots, u_{*m}$  は



$$\frac{1}{n}(\mathbf{UG}^T)^T \mathbf{UG}^T = \mathbf{G} \frac{\mathbf{U}^T \mathbf{U}}{n} \mathbf{G}^T = \mathbf{G} \mathbf{I}_m \mathbf{G}^T = \mathbf{I}_m \quad (17)$$

を満たすので、標準化され互いに無相関な共通因子と見なせる。また、最小二乗基準(16)は

$$2 \sum_{j < k} (\mathbf{b}_j^T \mathbf{b}_k - r_{jk})^2 = 2 \sum_{j < k} (\mathbf{b}_j^T \mathbf{G}^T \mathbf{G} \mathbf{b}_k - r_{jk})^2 = 2 \sum_{j < k} (\{\mathbf{G} \mathbf{b}_j\}^T \{\mathbf{G} \mathbf{b}_k\} - r_{jk})^2$$

と書けるので、最小二乗基準(16)を最小化する解  $\hat{\mathbf{b}}_j$  ( $j=1, \dots, p$ ) に対して直交行列  $\mathbf{G}$  で変換した

$$\mathbf{G} \hat{\mathbf{b}}_1, \dots, \mathbf{G} \hat{\mathbf{b}}_p \quad (18)$$

も最小二乗解となる。対応する共通因子  $\mathbf{u}_{*1}, \mathbf{u}_{*2}, \dots, \mathbf{u}_{*m}$  は互いに直交するので、(18)は直交解でもある。このように解が一意に定まらないことは解の不定性とよばれる。

因子分析では、この不定性を逆手に取って解釈がしやすい解を探索する。南風原 (2002) によると、因子の解釈の容易性の条件として、単純構造という基準がある。この基準は、

- 各因子には比較的少数の観測変数のみが高い負荷をもつ
- 各観測変数は1つの因子にだけ高い負荷をもつ

というものである。直交解の中で、単純構造を目指す解として、バリマックス法が知られている。共通因子  $u_\ell$  について、 $p$  個の観測変数の負荷の2乗  $b_{1\ell}^2, \dots, b_{p\ell}^2$  の平均を  $b_\ell^2$  とし、これらの分散

$$V_\ell = \frac{1}{p} \sum_{j=1}^p (b_{j\ell}^2 - b_\ell^2)^2$$

を考える。これを  $m$  個の因子すべてにわたって合計した値

$$V = \sum_{\ell=1}^m V_\ell$$

をバリマックス基準と定義する。初期解  $(\hat{\mathbf{b}}_1, \dots, \hat{\mathbf{b}}_p)$  をもとにして、この基準を最大にする  $(\mathbf{G} \hat{\mathbf{b}}_1, \dots, \mathbf{G} \hat{\mathbf{b}}_p)$  がバリマックス解である。図2はバリマックス解による可視化である。図1と図2はいずれも、因子負荷の絶対値が0.25を下回る場合に矢印を繋いでいない。バリマックス解が単純構造を目指す解であることが分かるだろう。

### 2.3 相関のある共通因子による $\hat{\mathbf{y}}_j$ の近似

(6)や(17)では、共通因子間の無相関性を仮定した。この節では、 $m$  個の共通因子  $\mathbf{v}_1, \dots, \mathbf{v}_m$  は標準化されており、 $\mathbf{v}_\ell$  と  $\mathbf{v}_q$  が相関を持ち、相関係数が  $\lambda_{\ell q}$  である状況を考える。つまり、 $m$  個の共通因子  $\mathbf{v}_1, \dots, \mathbf{v}_m$  の相関係数行列は

$$\frac{1}{n} \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_m^T \end{pmatrix} (\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_m) = \frac{\mathbf{V}^T \mathbf{V}}{n} = \begin{pmatrix} 1 & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & 1 & \cdots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1} & \lambda_{m2} & \cdots & 1 \end{pmatrix} = \mathbf{\Lambda} \quad (19)$$

となる。以下では、共通因子  $\mathbf{v}_1, \dots, \mathbf{v}_m$  の相関係数行列が(19)である場合1に、2.1節で与えられた概念を整理する。

相関を持つ共通因子  $\mathbf{v}_1, \dots, \mathbf{v}_m$  を用いて、(4)のように、第  $j$  特性の標準化ベクトル  $\hat{\mathbf{y}}_j$  ( $j=1, \dots, p$ ) を潜在変数による線形和

$$\mathbf{z}_j = c_{j1}\mathbf{v}_1 + \dots + c_{jm}\mathbf{v}_m + a_j\mathbf{e}_j = \mathbf{V}\mathbf{c}_j + a_j\mathbf{e}_j \quad (20)$$

で近似する。ただし、

$$\mathbf{c}_j = (c_{j1} \ c_{j2} \ \dots \ c_{jm})^T \in \mathbb{R}^m$$

である。(20)における係数  $c_{j\ell}$  ( $j=1, \dots, p, \ell=1, \dots, m$ ) は因子負荷とよばれ、因子負荷  $c_{j\ell}$  を  $(j, \ell)$  要素とする  $p \times m$  行列  $(\mathbf{c}_1 \ \mathbf{c}_2 \ \dots \ \mathbf{c}_p)^T$  は因子負荷行列とよばれる。また、独自因子  $\mathbf{e}_j$  ( $j=1, \dots, p$ ) は(7)のように互いに無相関な標準化ベクトルであり、(8)と同様に共通因子  $\mathbf{v}_1, \dots, \mathbf{v}_m$  との無相関性

$$\mathbf{e}_j^T \mathbf{v}_\ell = 0, \text{ for } j=1, \dots, p, \ell=1, \dots, m \quad (21)$$

を仮定する。

$\mathbf{z}_j$  は標準化ベクトルの線形和なので、 $\mathbf{1}_n^T \mathbf{z}_j = 0$  を満たす。また、(7), (19), (20), (21)より

$$\frac{\|\mathbf{z}_j\|^2}{n} = \frac{\|\mathbf{V}\mathbf{c}_j + a_j\mathbf{e}_j\|^2}{n} = \mathbf{c}_j^T \mathbf{\Lambda} \mathbf{c}_j + a_j^2 \quad (22)$$

を満たす。従って以下では

$$\mathbf{c}_j^T \mathbf{\Lambda} \mathbf{c}_j + a_j^2 = 1 \quad j=1, \dots, p \quad (23)$$

という制約を課して、 $\mathbf{z}_j$  を標準化ベクトルとして扱う。また  $\mathbf{z}_j$  と  $\mathbf{z}_k$  の相関係数 ( $j \neq k$ ) は、(22)と同様に

$$\frac{\mathbf{z}_j^T \mathbf{z}_k}{n} = \frac{1}{n} (\mathbf{c}_j^T \mathbf{V}^T + a_j \mathbf{e}_j^T) (\mathbf{V}\mathbf{c}_k + a_k \mathbf{e}_k) = \mathbf{c}_j^T \mathbf{\Lambda} \mathbf{c}_k$$

となり、 $\mathbf{z}_1, \dots, \mathbf{z}_p$  の相関係数行列は、

$$\frac{1}{n} \begin{pmatrix} \mathbf{z}_1^T \\ \mathbf{z}_2^T \\ \vdots \\ \mathbf{z}_p^T \end{pmatrix} (\mathbf{z}_1 \ \mathbf{z}_2 \ \dots \ \mathbf{z}_p) = \begin{pmatrix} 1 & \mathbf{c}_1^T \mathbf{\Lambda} \mathbf{c}_2 & \dots & \mathbf{c}_1^T \mathbf{\Lambda} \mathbf{c}_p \\ \mathbf{c}_2^T \mathbf{\Lambda} \mathbf{c}_1 & 1 & \dots & \mathbf{c}_2^T \mathbf{\Lambda} \mathbf{c}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_p^T \mathbf{\Lambda} \mathbf{c}_1 & \mathbf{c}_p^T \mathbf{\Lambda} \mathbf{c}_2 & \dots & 1 \end{pmatrix} \quad (24)$$

で与えられる。特に(24)の  $(j, j)$  対角成分の関係は

$$\frac{\|\mathbf{z}_j\|^2}{n} = \mathbf{c}_j^T \mathbf{\Lambda} \mathbf{c}_j + a_j^2 = 1 \quad (25)$$

である。(25)において、 $\mathbf{c}_j^T \mathbf{\Lambda} \mathbf{c}_j$  は  $\mathbf{z}_j$  のうち共通因子で説明できる割合で  $\mathbf{z}_j$  の共通性とよばれる。一方  $a_j^2$  は共通因子で説明できない割合で  $\mathbf{z}_j$  の独自性とよばれる。

$\mathbf{z}_j$  と共通因子  $\mathbf{v}_\ell$  の相関係数は、(19), (20), (21)より

$$\frac{\mathbf{z}_j^T \mathbf{v}_\ell}{n} = \sum_{q=1}^m c_{jq} \frac{\mathbf{v}_q^T \mathbf{v}_\ell}{n} + a_j \mathbf{e}_j^T \mathbf{v}_\ell = \sum_{q=1}^m c_{jq} \lambda_{q\ell} \quad (26)$$

であり因子構造とよばれる。(26)で与えられる因子構造  $\sum_{q=1}^m c_{jq} \lambda_{q\ell}$  を  $(j, \ell)$  要素とする  $p \times m$  行列  $(\mathbf{c}_1 \mathbf{c}_2 \dots \mathbf{c}_p)^T \mathbf{\Lambda}$  は因子構造行列とよばれる。共通因子の直交性のもとでは、(14)のように因子構造は因子負荷に一致するが、共通因子間に相関がある場合には異なることが分かる。また共通因子  $\mathbf{v}_\ell$  に関して、 $\mathbf{z}_1, \dots, \mathbf{z}_p$  との因子構造の二乗和

$$\sum_{j=1}^p \left\{ \sum_{q=1}^m c_{jq} \lambda_{q\ell} \right\}^2 \quad (27)$$

は共通因子  $\mathbf{v}_\ell$  の寄与とよばれる。この値が大きい因子ほど、潜在変数たち  $\mathbf{z}_1, \dots, \mathbf{z}_p$  と高い相関を持つことを意味する。

線形和(20)の共通因子部分の係数として定義された因子負荷、及び共通因子間の相関係数行列(19)の推定は、2節の(16)と同様である。つまり近似

$$\hat{\mathbf{y}}_j \approx \mathbf{z}_j = c_{j1} \mathbf{v}_1 + \dots + c_{jm} \mathbf{v}_m + a_j \mathbf{e}_j$$

の良さを、相関係数行列間の近さ(データから計算される相関係数行列(3)と潜在変数の線形和で表現された近似による相関係数行列(24)の差のフロベニウスノルム)

$$\sum_{j \neq k} (\mathbf{c}_j^T \mathbf{\Lambda} \mathbf{c}_k - r_{jk})^2 = 2 \sum_{j < k} (\mathbf{c}_j^T \mathbf{\Lambda} \mathbf{c}_k - r_{jk})^2 \quad (28)$$

で評価して最小化(ただし、 $\mathbf{c}_j^T \mathbf{\Lambda} \mathbf{c}_j \leq 1$  for  $j=1, \dots, p$  と制約)する。これにより、因子負荷  $c_{j\ell}$  ( $j=1, \dots, p, \ell=1, \dots, m$ ) と共通因子間の相関係数行列  $\mathbf{\Lambda}$  が推定される。特に因子負荷  $c_{j\ell}$  の推定値は斜交解とよばれる。斜交解にも不定性があり、単純構造を目指す様々な方法が知られている。

(25)で定義された共通性、独自性、(26)で定義された因子構造、(27)で定義された寄与は、全て相関係数行列と因子負荷の関数である。推定された因子負荷、相関係数行列を代入することで、共通性、独自性、因子構造、寄与が推定出来る。

### 3 確認的アプローチ

確認的アプローチは、因子数  $m$  および因子負荷について、一部の因子負荷  $b_{j\ell}$  を 0 に固定するなどの仮説的なモデルを用意し、そのモデルをデータによって検証することを目的とする。確認的アプローチでは、共通因子が互いに相関を持つケースが想定されるので、以下の記法は2.3節による。簡単のために  $p=8, m=2$  とし、 $\mathbf{y}_1, \dots, \mathbf{y}_4$  が共通因子 1  $\mathbf{v}_1$  に紐付けられ、 $\mathbf{y}_5, \dots, \mathbf{y}_8$  が共通因子 2  $\mathbf{v}_2$  に紐付けられるとする。このとき  $\mathbf{y}_1, \dots, \mathbf{y}_8$  を近似する  $\mathbf{z}_1, \dots, \mathbf{z}_8$  は、(20)に従って、

$$\begin{aligned} \mathbf{z}_j &= c_{j1} \mathbf{v}_1 + a_j \mathbf{e}_j = c_{j1} \mathbf{v}_1 + 0 \mathbf{v}_2 + a_j \mathbf{e}_j \quad (j=1, \dots, 4) \\ \mathbf{z}_j &= c_{j2} \mathbf{v}_2 + a_j \mathbf{e}_j = 0 \mathbf{v}_1 + c_{j2} \mathbf{v}_2 + a_j \mathbf{e}_j \quad (j=5, \dots, 8) \end{aligned}$$

と書ける。そして対象関数(28)、つまり

$$\sum_{j \neq k} (\mathbf{c}_j^T \Lambda \mathbf{c}_k - r_{jk})^2 \tag{29}$$

を

$$c_{j1} = 0 \quad (j=1, \dots, 4), \quad c_{j2} = 0 \quad (j=5, \dots, 8),$$

及び、(23)つまり  $\mathbf{c}_j^T \Lambda \mathbf{c}_j \leq 1$  for  $j=1, \dots, 8$  の制約のもとに、最小二乗解を求める。

表2 共分散構造分析のための相関係数行列

y1	1												
y2	0.16	1											
y3	0.3	0.34	1										
y4	0.46	0.4	0.37	1									
y5	0.3	0.4	0.55	0.3	1								
y6	0.15	0.32	0.48	0.22	0.71	1							
y7	0.13	0.4	0.47	0.26	0.62	0.32	1						
y8	0.18	0.37	0.57	0.26	0.78	0.77	0.72	1					
y9	0.25	0.28	0.32	0.16	0.36	0.3	0.26	0.28	1				
y10	0.37	0.1	0.41	0.24	0.29	0.21	0.07	0.14	0.3	1			
y11	0.16	0.29	0.39	0.23	0.47	0.35	0.2	0.32	0.29	0.47	1		
y12	0.21	-0.01	0.37	0.22	0.34	0.2	0.15	0.19	0.42	0.35	0.38	1	
	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12	

#### 4 共分散構造分析

本節では  $y_1, \dots, y_{12}$  の相関係数行列 (表2, 南風原 (2002) の表10.8) 及びパス図 (図3, 南風原 (2002) の図10.6) を念頭に置いて共分散構造分析を説明する。因子分析と同様に、 $y_1, \dots, y_{12}$  の標準化ベクトル  $\tilde{\mathbf{y}}_j$  ( $j=1, \dots, 12$ ) に対して近似

$$\tilde{\mathbf{y}}_j \approx \mathbf{z}_j$$

を満たす潜在変数たちの線形和

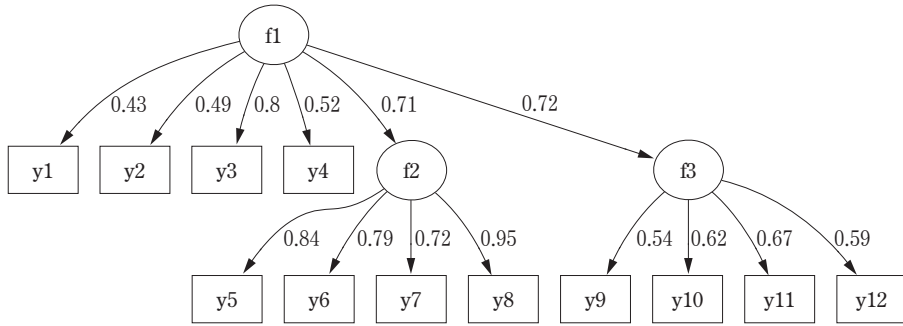
$$\mathbf{z}_j \propto \begin{cases} \beta_{j1} \mathbf{f}_1 + \alpha_j \mathbf{e}_j & j=1, \dots, 4 \\ \beta_{j2} \mathbf{f}_2 + \alpha_j \mathbf{e}_j & j=5, \dots, 8 \\ \beta_{j3} \mathbf{f}_3 + \alpha_j \mathbf{e}_j & j=9, \dots, 12 \end{cases} \tag{30}$$

を構成したい。ここで、 $\mathbf{f}_l$  と  $\mathbf{e}_j$  ( $j=1, \dots, 12$ ) は標準化ベクトルであり、 $\mathbf{f}_2$  と  $\mathbf{f}_3$  は中心化ベクトル ( $n$  個の成分の算術平均が0ではあるが、標本分散が1とは限らない) である。また、 $\mathbf{f}_l$  ( $l=1, 2, 3$ ) と  $\mathbf{e}_j$  ( $j=1, \dots, 12$ )、 $\mathbf{e}_j$  と  $\mathbf{e}_k$  ( $j \neq k$ ) は互いに無相関であり

$$\begin{aligned} \mathbf{f}_l^T \mathbf{e}_j &= 0 \quad \text{for } l=1, 2, 3 \\ \mathbf{e}_j^T \mathbf{e}_k &= 0 \quad j \neq k \end{aligned} \tag{31}$$

を満たす。この段階で  $\mathbf{z}_j$  ( $j=1, \dots, 12$ ) は標準化ベクトルではなく、中心化ベクトルであり、後で標準化する。その意味で(30)式では比例記号  $\propto$  としている。

図3 共分散構造分析のパス図



(30)の表現の一意性について、共分散構造分析では2種類の制約が使われる。1つは、

$$\beta_{11} = \beta_{22} = \beta_{33} = 1,$$

$$\alpha_1 = \dots = \alpha_4, \alpha_5 = \dots = \alpha_8, \alpha_9 = \dots = \alpha_{12}$$

であり、もう一つは

$$\alpha_j = 1 \quad \text{for } j=1, \dots, 12$$

である。ここでは南風原 (2002) に従い、後者を採用する。このとき  $z_1, \dots, z_{12}$  の標準化は

$$z_j = \begin{cases} \frac{\beta_{j1}f_1 + e_j}{\sqrt{\beta_{j1}^2 + 1}} & j=1, \dots, 4 \\ \frac{\beta_{j2}f_2 + e_j}{\sqrt{\beta_{j2}^2 \|f_2\|^2/n + 1}} & j=5, \dots, 8 \\ \frac{\beta_{j3}f_3 + e_j}{\sqrt{\beta_{j3}^2 \|f_3\|^2/n + 1}} & j=9, \dots, 12 \end{cases} \quad (32)$$

で与えられる。

さて共分散構造分析においては、共通因子間に回帰関係を想定する。南風原 (2002) に従い、共通因子  $f_2, f_3$  が共通因子  $f_1$  を説明変数として

$$f_2 = \gamma_{21}f_1 + \nu_2 \delta_2 \quad (33)$$

$$f_3 = \gamma_{31}f_1 + \nu_3 \delta_3$$

のような回帰関係を持つと想定する。(33)で、 $\delta_2$  と  $\delta_3$  は互いに無相関な標準化ベクトルで、それぞれ  $e_j$  ( $j=1, \dots, 12$ ) 及び  $f_1$  とも無相関であると仮定する。つまり  $f_1, \delta_2, \delta_3$  は

$$\begin{aligned} \delta_2^\top \delta_3 &= 0 \\ \delta_2^\top f_1 &= \delta_3^\top f_1 = 0 \end{aligned} \quad (34)$$

を満たす。また、(31)と(34)より

$$\delta_2^\top e_j = \delta_3^\top e_j = 0 \quad j=1, \dots, 12$$

が従う。さらに(33)の表現の一意性を確保するために、南風原 (2002) に従い、 $\nu_2 = \nu_3 = 1$  を採用する。つまり

$$\begin{aligned} \mathbf{f}_2 &= \gamma_{21} \mathbf{f}_1 + \boldsymbol{\delta}_2 \\ \mathbf{f}_3 &= \gamma_{31} \mathbf{f}_1 + \boldsymbol{\delta}_3 \end{aligned} \quad (35)$$

である。(35)に対して、 $\mathbf{f}_1$ が標準化ベクトルであること、及び(34)より

$$\frac{\|\mathbf{f}_2\|^2}{n} = \frac{\|\gamma_{21} \mathbf{f}_1 + \boldsymbol{\delta}_2\|^2}{n} = \gamma_{21}^2 \frac{\|\mathbf{f}_1\|^2}{n} + \frac{\|\boldsymbol{\delta}_2\|^2}{n} = \gamma_{21}^2 + 1 \quad (36)$$

$$\frac{\|\mathbf{f}_3\|^2}{n} = \frac{\|\gamma_{31} \mathbf{f}_1 + \boldsymbol{\delta}_3\|^2}{n} = \gamma_{31}^2 \frac{\|\mathbf{f}_1\|^2}{n} + \frac{\|\boldsymbol{\delta}_3\|^2}{n} = \gamma_{31}^2 + 1 \quad (37)$$

が従う。

(36)および(37)より、 $\mathbf{f}_2$ と $\mathbf{f}_3$ を標準化して

$$\mathbf{v}_2 = \frac{\mathbf{f}_2}{\sqrt{\gamma_{21}^2 + 1}}, \quad \mathbf{v}_3 = \frac{\mathbf{f}_3}{\sqrt{\gamma_{31}^2 + 1}}$$

とし、また記法の統一のため $\mathbf{v}_1 = \mathbf{f}_1$ としよう。このとき、 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ の間の相関係数は

$$\frac{1}{n} \mathbf{v}_1^\top \mathbf{v}_2 = \frac{\gamma_{21}}{\sqrt{\gamma_{21}^2 + 1}} = \lambda_{12} \quad (38)$$

$$\frac{1}{n} \mathbf{v}_1^\top \mathbf{v}_3 = \frac{\gamma_{31}}{\sqrt{\gamma_{31}^2 + 1}} = \lambda_{13} \quad (39)$$

$$\frac{1}{n} \mathbf{v}_2^\top \mathbf{v}_3 = \frac{\gamma_{21} \gamma_{31}}{\sqrt{\gamma_{21}^2 + 1} \sqrt{\gamma_{31}^2 + 1}} = \lambda_{12} \lambda_{13} \quad (40)$$

であり、従って $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ の相関係数行列 $\Lambda$ は、

$$\Lambda = \begin{pmatrix} 1 & \lambda_{12} & \lambda_{13} \\ \lambda_{12} & 1 & \lambda_{12} \lambda_{13} \\ \lambda_{13} & \lambda_{12} \lambda_{13} & 1 \end{pmatrix} \quad (41)$$

のように2つのパラメータ $\lambda_{12}, \lambda_{13}$ で表現される。

(32)における $\mathbf{z}_j$ を $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ の線形和で表現するにあたり、

$$c_{j1} = \frac{\beta_{j1}}{\sqrt{\beta_{j1}^2 + 1}}, \quad a_j = \frac{1}{\sqrt{\beta_{j1}^2 + 1}}, \quad j=1, \dots, 4 \quad (42)$$

$$c_{j2} = \frac{\beta_{j2} \sqrt{\gamma_{21}^2 + 1}}{\sqrt{\beta_{j2}^2 (\gamma_{21}^2 + 1) + 1}}, \quad a_j = \frac{1}{\sqrt{\beta_{j2}^2 (\gamma_{21}^2 + 1) + 1}}, \quad j=5, \dots, 8 \quad (43)$$

$$c_{j3} = \frac{\beta_{j3} \sqrt{\gamma_{31}^2 + 1}}{\sqrt{\beta_{j3}^2 (\gamma_{31}^2 + 1) + 1}}, \quad a_j = \frac{1}{\sqrt{\beta_{j3}^2 (\gamma_{31}^2 + 1) + 1}}, \quad j=9, \dots, 12 \quad (44)$$

とすると、(32)の標準化ベクトルは

$$\mathbf{z}_j = \begin{cases} c_{j1} \mathbf{v}_1 + a_j \mathbf{e}_j & j=1, \dots, 4 \\ c_{j2} \mathbf{v}_2 + a_j \mathbf{e}_j & j=5, \dots, 8 \\ c_{j3} \mathbf{v}_3 + a_j \mathbf{e}_j & j=9, \dots, 12 \end{cases}$$

と表現できる。また  $v_1, v_2, v_3$  の相関係数行列は(41)で与えられるので、確認的因子分析の特別な場合（共通因子間の相関構造に制約がある）と見なせる。従って、3節の最小二乗法(29)によって、パラメータ

$$\begin{cases} c_{j1} & j=1, \dots, 4, \\ c_{j2} & j=5, \dots, 8 \\ c_{j3} & j=9, \dots, 12 \\ a_j & j=1, \dots, 12 \\ \lambda_{12}, \lambda_{13} \end{cases}$$

が推定される。そして(38), (39), (40), (41), (42), (43), (44)の関係から変換されて、共分散構造モデルのパラメータ

$$\begin{cases} \beta_{j1} & j=1, \dots, 4, \\ \beta_{j2} & j=5, \dots, 8 \\ \beta_{j3} & j=9, \dots, 12 \\ \gamma_{21}, \gamma_{31} \end{cases}$$

の推定値が得られる。

#### 参 考 文 献

- 市川雅教『因子分析』（朝倉書店，2010）  
 南風原朝和『心理統計学の基礎』（有斐閣，2002）