データ解析

http://coconut.sys.eng.shizuoka.ac.jp/data/

静岡大学工学部
安藤和敏

2006.01.11

第4章 Excelで学ぶ因子分析

4-1　1因子モデルから学ぶ因子分析の考え方

因子分析とは

我々は，複雑な現象を単純な原因（すなわち因子）で理解することがよくある．

例）「彼は理系の才能があるので理科が得意であるが文型の才能がないので国語が苦手である．」
複雑な人間の能力を「理系的才能」と「文系的才能」という2つの因子で単純に説明．

例）「K君はO型だからいいかげんである．Lさんは，A型だから真面目である．」
複雑な人間の性格を「血液型」という1つの因子で単純に説明．

因子分析とは

因子分析とは，複雑な現象を単純な因子で説明するための，統計学的な手法である．

因子分析のデータ（変数が3個の場合）

No	国語 (x)	英語 (y)	数学 (z)
1	x_1	y_1	z_1
2	x_2	y_2	z_2
⋮	⋮	⋮	⋮
i	x_i	y_i	z_i
⋮	⋮	⋮	⋮
n	x_n	y_n	z_n

この多変量データに対して，「学力」という1つの共通因子を仮定して，因子分析を行う．

この多変量データは標準化されていると仮定する．

標準化（第3回のスライドから）

$$x'_i = \frac{x_i - \bar{x}}{s_x} \qquad (i = 1, \cdots, n)$$

標準化された変数の平均は0，分散は1になる．（証明せよ．）

$$\overline{x'} = 0 \,,\; s_{x'} = 1$$

因子分析の記法

「学力」という共通因子を変数Fで表してみる．

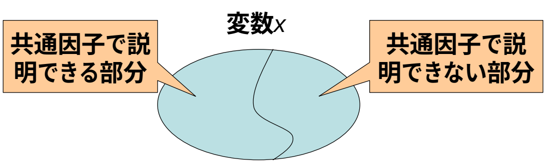
No	因子（F）	国語（x）	英語（y）	数学（z）
1	F_1	x_1	y_1	z_1
2	F_2	x_2	y_2	z_2
⋮	⋮	⋮	⋮	⋮
n	F_n	x_n	y_n	z_n

学籍番号iの学生の学力はF_iである．これをi番目の学生の**因子得点**と呼ぶ．

共通因子と独自因子

各科目の得点が共通因子Fだけで説明できることはありえない．

変数 x は，共通因子で説明できる部分と，共通因子で説明できない部分に分けることができると考える．



因子分析のモデル

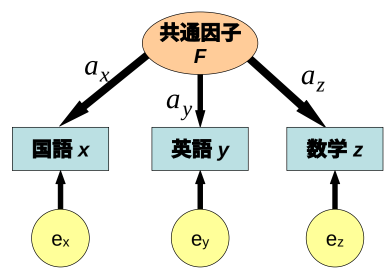
$$x = a_x F + e_x \cdots\cdots (1)$$

因子負荷量　共通因子　独自因子

因子分析のモデル（3変数の場合）

$$x = a_x F + e_x \quad \cdots\cdots \quad (1)$$
$$y = a_y F + e_y \quad \cdots\cdots \quad (2)$$
$$z = a_z F + e_z \quad \cdots\cdots \quad (3)$$

因子分析法のパス図



因子分析のデータ (変数が3個の場合)

No	F	x	y	z
1	F_1	$x_1 = a_x F_1 + e_{x1}$	$y_1 = a_y F_1 + e_{y1}$	$z_1 = a_z F_1 + e_{z1}$
2	F_2	$x_2 = a_x F_2 + e_{x2}$	$y_2 = a_y F_2 + e_{y2}$	$z_2 = a_z F_2 + e_{z2}$
⋮	⋮	⋮	⋮	⋮
i	F_i	$x_i = a_x F_i + e_{xi}$	$y_i = a_y F_i + e_{yi}$	$z_i = a_z F_i + e_{zi}$
⋮	⋮	⋮	⋮	⋮
n	F_n	$x_n = a_x F_n + e_{xn}$	$y_n = a_y F_n + e_{yn}$	$z_n = a_z F_n + e_{zn}$

因子分析の基本方程式 (⇒WB)

$$\left.\begin{aligned}
s_x^2 &= a_x^2 s_F^2 + 2a_x s_{Fe_x} + s_{e_x}^2 \\
s_y^2 &= a_y^2 s_F^2 + 2a_y s_{Fe_y} + s_{e_y}^2 \\
s_z^2 &= a_z^2 s_F^2 + 2a_z s_{Fe_z} + s_{e_z}^2 \\
s_{xy} &= a_x a_y s_F^2 + a_x s_{Fe_y} + a_y s_{Fe_x} + s_{e_x e_y} \\
s_{yz} &= a_y a_z s_F^2 + a_y s_{Fe_z} + a_z s_{Fe_y} + s_{e_y e_z} \\
s_{zx} &= a_z a_x s_F^2 + a_z s_{Fe_x} + a_x s_{Fe_z} + s_{e_z e_x}
\end{aligned}\right\}\cdots(5)$$

重要な仮定

「共通因子と独自因子は，互いに無相関 (⇒p. 19) である」と仮定する．

つまり，以下の式を仮定する．

$$s_{Fe_x} = s_{Fe_y} = s_{Fe_z} = 0,$$

$$s_{e_x e_y} = s_{e_y e_z} = s_{e_z e_x} = 0 \cdots\cdots(6)$$

因子分析の基本方程式

$$\left.\begin{aligned}
s_x^2 &= a_x^2 s_F^2 + 2a_x \boxed{s_{Fe_x}} + s_{e_x}^2 \\
s_y^2 &= a_y^2 s_F^2 + 2a_y \boxed{s_{Fe_y}} + s_{e_y}^2 \\
s_z^2 &= a_z^2 s_F^2 + 2a_z \boxed{s_{Fe_z}} + s_{e_z}^2 \\
s_{xy} &= a_x a_y s_F^2 + a_x \boxed{s_{Fe_y}} + a_y \boxed{s_{Fe_x}} + \boxed{s_{e_x e_y}} \\
s_{yz} &= a_y a_z s_F^2 + a_y \boxed{s_{Fe_z}} + a_z \boxed{s_{Fe_y}} + \boxed{s_{e_y e_z}} \\
s_{zx} &= a_z a_x s_F^2 + a_z \boxed{s_{Fe_x}} + a_x \boxed{s_{Fe_z}} + \boxed{s_{e_z e_x}}
\end{aligned}\right\}\cdots(5)$$

因子分析の基本方程式

$$\left.\begin{aligned}
s_x^2 &= a_x^2 s_F^2 + s_{e_x}^2 \\
s_y^2 &= a_y^2 s_F^2 + s_{e_y}^2 \\
s_z^2 &= a_z^2 s_F^2 + s_{e_z}^2 \\
s_{xy} &= a_x a_y s_F^2 \\
s_{yz} &= a_y a_z s_F^2 \\
s_{zx} &= a_z a_x s_F^2
\end{aligned}\right\}\cdots\cdots(7)$$

最初の仮定

x, y, z は標準化されていると仮定した．

つまり，以下を仮定する．

$$s_x^2 = s_y^2 = s_z^2 = 1,$$

$$s_{xy} = r_{xy}, s_{yz} = r_{yz}, s_{zx} = r_{zx}$$

さらに，F も標準化されていると仮定する．

$$s_F^2 = 1$$

因子分析の基本方程式

$$
\left.\begin{aligned}
\boxed{s_x^2} &= a_x^2\boxed{s_F^2}+s_{e_x}^2\\
\boxed{s_y^2} &= a_y^2\boxed{s_F^2}+s_{e_y}^2\\
\boxed{s_z^2} &= a_z^2\boxed{s_F^2}+s_{e_z}^2\\
r_{xy}=s_{xy} &= a_x a_y\boxed{s_F^2}\\
r_{yz}=s_{yz} &= a_y a_z\boxed{s_F^2}\\
r_{zx}=s_{zx} &= a_z a_x\boxed{s_F^2}
\end{aligned}\right\}\cdots\cdots(7)
$$

$\boxed{1}$

因子分析の基本方程式 (3変数の場合)

$$
\left.\begin{aligned}
1 &= a_x^2+s_{e_x}^2\\
1 &= a_y^2+s_{e_y}^2\\
1 &= a_z^2+s_{e_z}^2\\
r_{xy} &= a_x a_y\\
r_{yz} &= a_y a_z\\
r_{zx} &= a_z a_x
\end{aligned}\right\}\cdots\cdots(8),(9)
$$

因子分析の基本方程式

$$
\begin{bmatrix} 1 & r_{xy} & r_{zx}\\ r_{xy} & 1 & r_{yz}\\ r_{zx} & r_{yz} & 1 \end{bmatrix}
=
\begin{bmatrix} a_x^2+s_{e_x}^2 & a_x a_y & a_x a_z\\ a_x a_y & a_y^2+s_{e_y}^2 & a_y a_z\\ a_x a_z & a_y a_z & a_z^2+s_{e_z}^2 \end{bmatrix}
$$

因子分析の基本方程式を解く

(9)式より, $r_{xy}\cdot r_{yz}\cdot r_{zx}=a_x^2\cdot a_y^2\cdot a_z^2$

$$
\therefore a_x\cdot a_y\cdot a_z=\sqrt{r_{xy}\cdot r_{yz}\cdot r_{zx}}\cdots\cdots(10)
$$

因子分析の基本方程式を解く

(10)式の両辺を(9)の各式で割ると,

$$
\left.\begin{aligned}
a_x &= \frac{\sqrt{r_{xy}\cdot r_{yz}\cdot r_{zx}}}{r_{yz}}\\
a_y &= \frac{\sqrt{r_{xy}\cdot r_{yz}\cdot r_{zx}}}{r_{zx}}\\
a_z &= \frac{\sqrt{r_{xy}\cdot r_{yz}\cdot r_{zx}}}{r_{xy}}
\end{aligned}\right\}\cdots\cdots(11)
$$

因子分析の基本方程式を解く

(11)の各式を(8)の各式に代入すると,

$$
\begin{aligned}
s_{e_x}^2 &= 1-a_x^2 = 1-\frac{r_{xy}\cdot r_{zx}}{r_{yz}}\\
s_{e_y}^2 &= 1-a_y^2 = 1-\frac{r_{xy}\cdot r_{yz}}{r_{zx}}\\
s_{e_z}^2 &= 1-a_z^2 = 1-\frac{r_{yz}\cdot r_{zx}}{r_{xy}}
\end{aligned}
$$

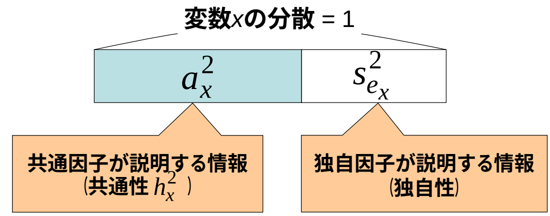
共通性と独自性

$$x = a_x F + e_x \cdots\cdots (1)$$

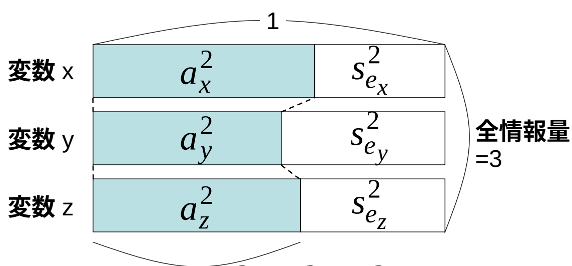
因子負荷量　共通因子　独自因子

共通性と独自性

$$s_x^2 = a_x^2 + s_{e_x}^2 = 1 \cdots\cdots (12)$$



総共通性と寄与率



総共通性 = $a_x^2 + a_y^2 + a_z^2 \cdots\cdots (13)$

総共通性と寄与率

$$\text{寄与率} = \frac{\text{総共通性}}{\text{全情報量}}$$

$$= \frac{a_x^2 + a_y^2 + a_z^2}{\text{変数の個数}}$$

Excelで学ぼう

ファイル：第4章/4_1

本日のまとめ

- 因子分析の考え方を理解した．
- 因子分析で用いられる言葉（因子負荷量，共通因子，独自因子，因子得点，寄与率など）の意味を理解した．
- 3変数の場合に，因子負荷量をExcelを用いて，計算する方法を理解した．