

Rによる因子分析：fa関数の実践と理論

— ml, uls, wls, gls, pa, varimax, promaxとは —

[本稿のWebページ](#)

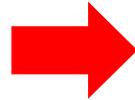
古橋 武

1. データの形式

	x_1	x_2	x_3	x_4	x_5
$i = 1$	数学	理科	英語	国語	社会
1	68	61	88	57	52
2	27	28	0	17	45
⋮	64	61	76	58	64
⋮	10	29	100	100	85
⋮	0	65	81	67	77
⋮	100	100	100	100	100
⋮	39	69	89	73	68
⋮	88	60	48	55	58
⋮	100	100	83	70	74
⋮	21	46	84	82	81
⋮	71	92	87	74	79
⋮	14	61	91	83	94
⋮	80	61	91	75	74
⋮	30	37	55	96	50
⋮	40	45	33	19	34
⋮	17	49	10	27	33
⋮	76	61	25	53	29
⋮	81	42	72	83	82
⋮	61	76	31	43	43
20	16	50	0	22	20

変数の数 $p = 5$

転置して
用いる.

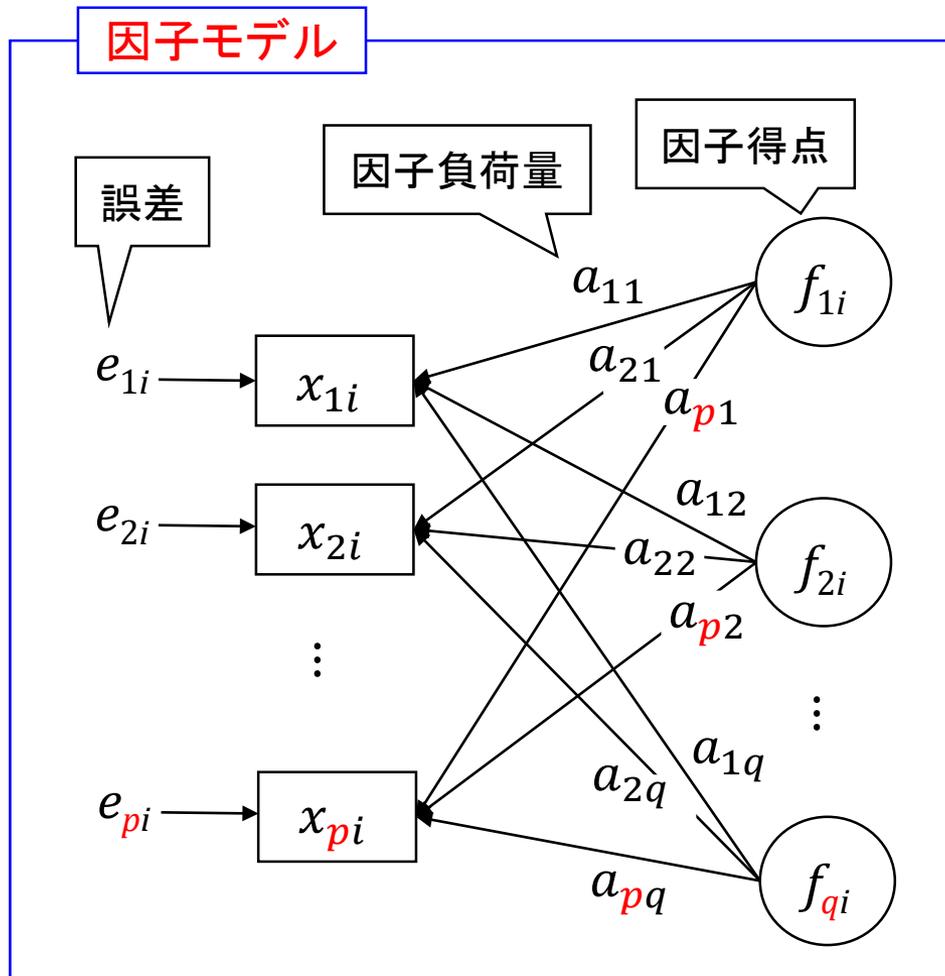


$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix}$$

$$= \begin{pmatrix} 68 & 27 & \cdots & 16 \\ 61 & 28 & \cdots & 50 \\ 88 & 0 & \cdots & 0 \\ 57 & 17 & \cdots & 22 \\ 52 & 45 & \cdots & 20 \end{pmatrix}$$

サンプルサイズ $n = 20$

2. 因子分析の目的 ➡ 因子モデルの同定



変数の数: p
 因子の数: q
 データのサイズ: n

$$\mathbf{x}_i = \mathbf{A}\mathbf{f}_i + \mathbf{e}_i: \text{因子モデル}$$

$$\mathbf{x}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{pi} \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix}$$

$$\mathbf{f}_i = \begin{pmatrix} f_{1i} \\ f_{2i} \\ \vdots \\ f_{qi} \end{pmatrix}, \mathbf{e}_i = \begin{pmatrix} e_{1i} \\ e_{2i} \\ \vdots \\ e_{pi} \end{pmatrix}$$

$$i = 1, 2, \dots, n$$

3. 因子モデル

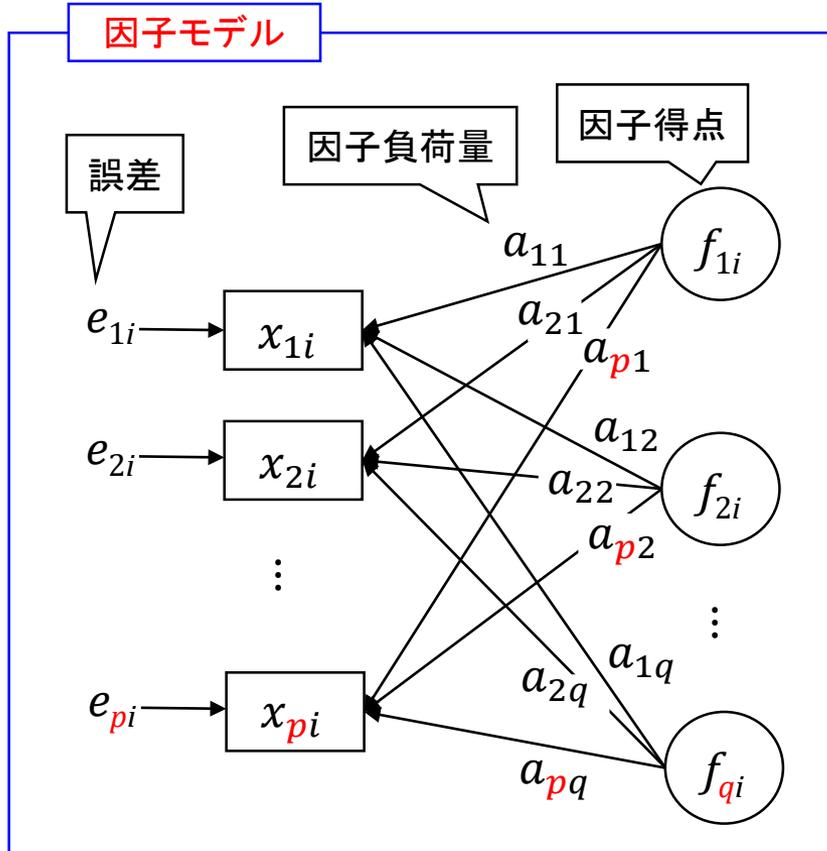
3.1 因子モデルの行列表現

$$\mathbf{X} = \mathbf{AF} + \mathbf{U}$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \cdots x_{1n} \\ x_{21} & x_{22} \cdots x_{2n} \\ \vdots & \vdots \ddots \vdots \\ x_{p1} & x_{p2} \cdots x_{pn} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \cdots a_{1q} \\ a_{21} & a_{22} \cdots a_{2q} \\ \vdots & \vdots \ddots \vdots \\ a_{p1} & a_{p2} \cdots a_{pq} \end{pmatrix}$$

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} \cdots f_{1n} \\ f_{21} & f_{22} \cdots f_{2n} \\ \vdots & \vdots \ddots \vdots \\ f_{q1} & f_{q2} \cdots f_{qn} \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} e_{11} & e_{12} \cdots e_{1n} \\ e_{21} & e_{22} \cdots e_{2n} \\ \vdots & \vdots \ddots \vdots \\ e_{p1} & e_{p2} \cdots e_{pn} \end{pmatrix}$$

3.2 因子モデルの仮定



変数の数: p

因子の数: q

データのサイズ: n

サンプルサイズが n で既知数が $p \times n$ 個
未知数が $p \times q + q \times n$ 個

⇒ 連立方程式は不定

そこで、以下の仮定を設ける。

① f_{ki} と e_{li} は無相関 $\sum_{i=1}^n f_{ki} e_{li} = 0$ ($k \neq l$)

② e_{ki} と e_{li} は無相関 $\sum_{i=1}^n e_{ki} e_{li} = 0$ ($k \neq l$)

③ f_{ki} と f_{li} は無相関 $\sum_{i=1}^n f_{ki} f_{li} = 0$ ($k \neq l$)

④ f_{ki} は正規化されている。

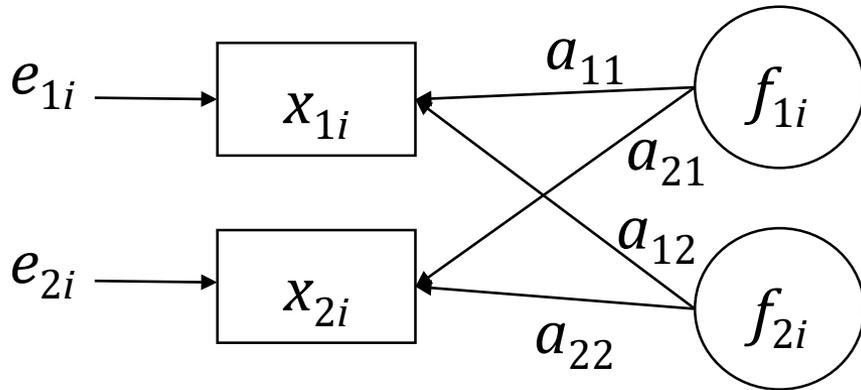
$$\bar{f}_{ki} = 0, \frac{1}{n} \sum_{i=1}^n f_{ki}^2 = 1$$

⑤ x_{ki} は正規化されている。

$$\bar{x}_k = 0, \frac{1}{n} \sum_{i=1}^n x_{ki}^2 = 1$$

3.3 仮定下でのモデルの単純化

簡単のため因子数2個とする.



サンプルサイズが n の時

$$\begin{cases} x_{11} = a_{11}f_{11} + a_{12}f_{21} + e_{11} \\ x_{21} = a_{21}f_{11} + a_{22}f_{21} + e_{21} \\ \\ x_{12} = a_{11}f_{12} + a_{12}f_{22} + e_{12} \\ x_{22} = a_{21}f_{12} + a_{22}f_{22} + e_{22} \\ \vdots \\ \vdots \\ \vdots \\ \\ x_{1n} = a_{11}f_{1n} + a_{12}f_{2n} + e_{1n} \\ x_{2n} = a_{21}f_{1n} + a_{22}f_{2n} + e_{2n} \end{cases}$$

3.3 仮定下でのモデルの単純化

$$\underbrace{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{pmatrix}}_{\mathbf{X}} = \begin{pmatrix} a_{11}f_{11} + a_{12}f_{21} & a_{11}f_{12} + a_{12}f_{22} & \cdots & a_{11}f_{1n} + a_{12}f_{2n} \\ a_{21}f_{11} + a_{22}f_{21} & a_{21}f_{12} + a_{22}f_{22} & \cdots & a_{21}f_{1n} + a_{22}f_{2n} \end{pmatrix} + \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \end{pmatrix}$$

$$\mathbf{X} = \underbrace{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \end{pmatrix}}_{\mathbf{F}} + \underbrace{\begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \end{pmatrix}}_{\mathbf{U}}$$

$$\mathbf{X} = \mathbf{A}\mathbf{F} + \mathbf{U}$$

$$\frac{1}{n} \mathbf{X}\mathbf{X}^t = \frac{1}{n} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{pmatrix} \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ \vdots & \vdots \\ x_{1n} & x_{2n} \end{pmatrix}$$

$$= \frac{1}{n} \begin{pmatrix} \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{1i}x_{2i} & \sum x_{2i}^2 \end{pmatrix}$$

x_{1i}, x_{2i} が規格化されているのでこれは相関係数となる

仮定より1

$$= \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} = \mathbf{S} \approx \mathbf{\Sigma}$$

真の相関行列

サンプルサイズ n が十分に大きければ成立する.

$$\begin{aligned} \frac{1}{n} \mathbf{X} \mathbf{X}^t &= \frac{1}{n} (\mathbf{A} \mathbf{F} + \mathbf{U})(\mathbf{A} \mathbf{F} + \mathbf{U})^t \\ &= \frac{1}{n} (\mathbf{A} \mathbf{F} + \mathbf{U})(\mathbf{F}^t \mathbf{A}^t + \mathbf{U}^t) \\ &= \frac{1}{n} \underbrace{\mathbf{A} \mathbf{F} \mathbf{F}^t \mathbf{A}^t}_{\text{}} + \frac{1}{n} \underbrace{\mathbf{A} \mathbf{F} \mathbf{U}^t}_{\text{}} + \frac{1}{n} \underbrace{\mathbf{U} \mathbf{F}^t \mathbf{A}^t}_{\text{}} + \frac{1}{n} \mathbf{U} \mathbf{U}^t \end{aligned}$$

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n f_{ki}^2 &= 1 \\ \sum_{i=1}^n f_{ki} f_{li} &= 0 \quad (k \neq l) \end{aligned}$$

$$\sum_{i=1}^n f_i e_{ji} = 0$$

$$\sum_{i=1}^n f_i e_{ji} = 0$$

$$= \mathbf{A} \mathbf{A}^t + \frac{1}{n} \mathbf{U} \mathbf{U}^t$$

$$= \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11} a_{21} + a_{12} a_{22} \\ a_{21} a_{11} + a_{22} a_{12} & a_{21}^2 + a_{22}^2 \end{pmatrix} + \frac{1}{n} \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{11} & e_{12} & \cdots & e_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1n} & e_{2n} & \cdots & e_{2n} \end{pmatrix}$$

$$= \mathbf{A} \mathbf{A}^t + \frac{1}{n} \begin{pmatrix} \sum e_{1i}^2 & \sum e_{1i} e_{2i} \\ \sum e_{1i} e_{2i} & \sum e_{2i}^2 \end{pmatrix}$$

仮定より 0

$$= \mathbf{A} \mathbf{A}^t + \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}$$

仮定より 0

簡単化された因子モデル

$$\underbrace{\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}}_{\mathbf{S}} = \underbrace{\begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 \end{pmatrix}}_{\mathbf{AA}^t} + \underbrace{\begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}}_{\boldsymbol{\psi}}$$

$$\mathbf{S} = \mathbf{AA}^t + \boldsymbol{\Psi}$$

d_i^2 : 独自性

$1 - d_i^2 = h_i^2$: 共通性

簡単化された因子モデル

$$\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} = \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 \end{pmatrix} + \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}$$

$S \approx \Sigma$

データから決まる

AA^t

一意には決められない

ψ

一意には決められない

$\frac{p(p+1)}{2}$ 個が既知

$p \times q$ 個が未知

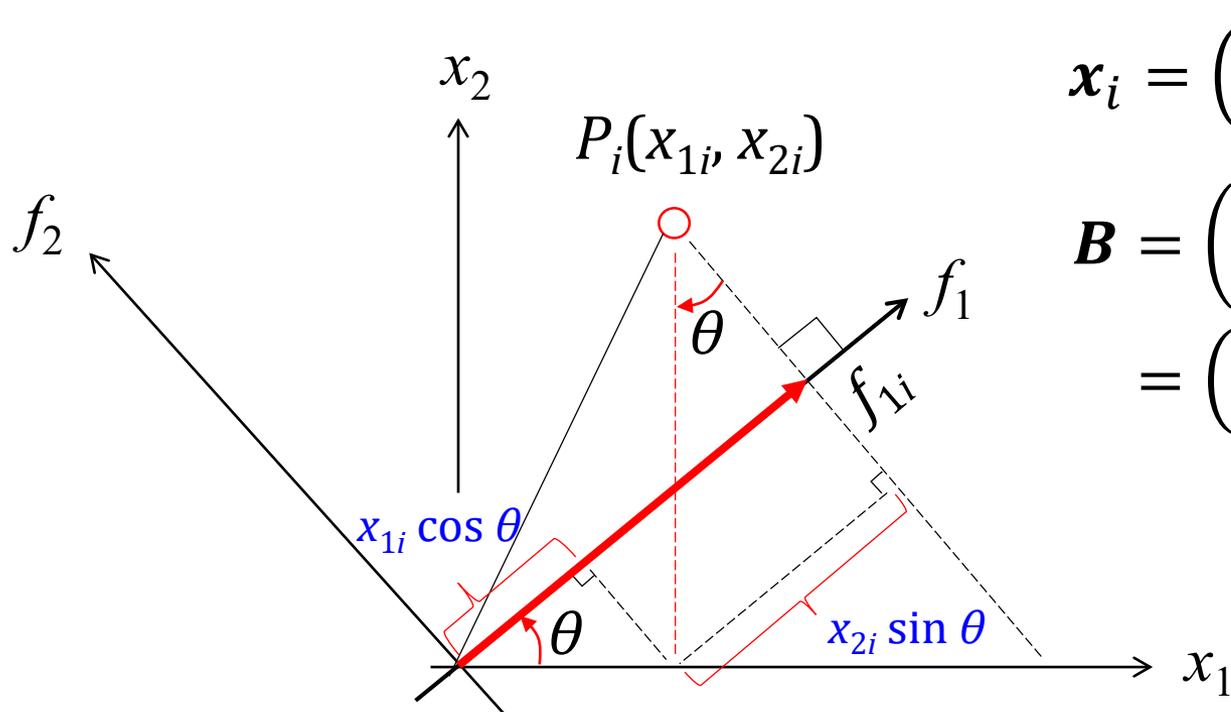
p 個が未知



基準を基に A, Ψ を決定する

4. 主成分分析との違いは？

主成分分析



f_1, f_2 は主成分軸

$$\mathbf{x}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix} \quad \mathbf{f}_i = \begin{pmatrix} f_{1i} \\ f_{2i} \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

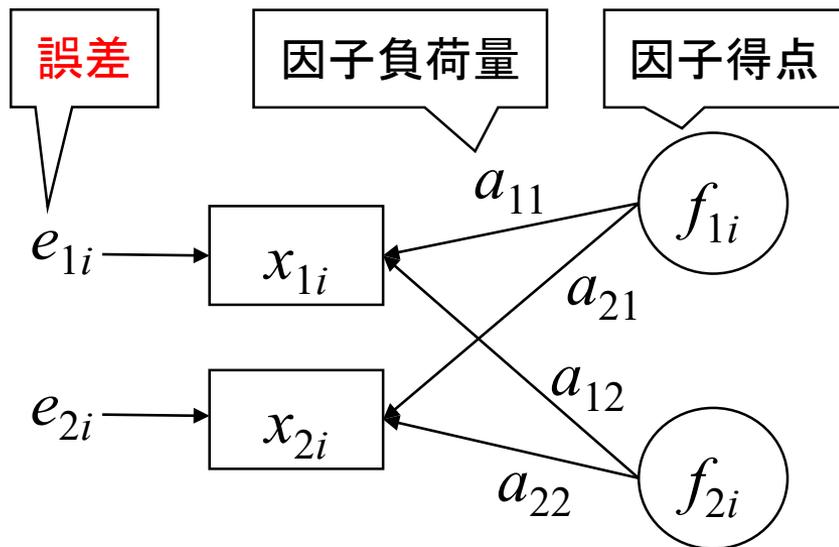
$$\mathbf{f}_i = \mathbf{B} \mathbf{x}_i$$

$$\mathbf{x}_i = \mathbf{B}^{-1} \mathbf{f}_i$$

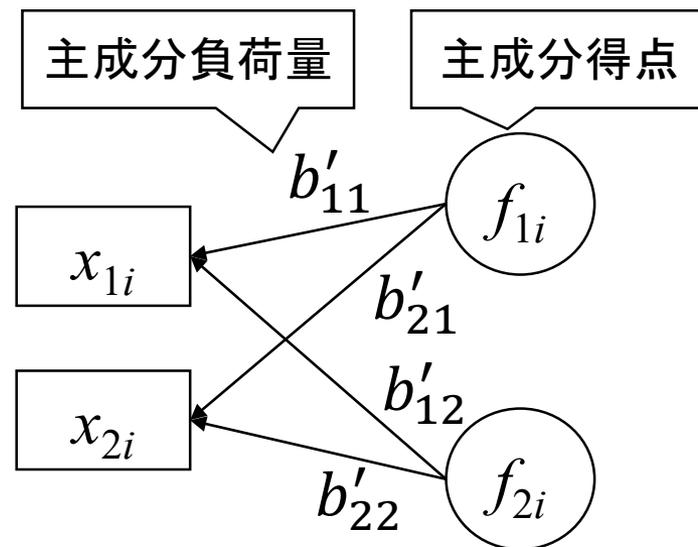
$$\mathbf{X} = \mathbf{B}^{-1} \mathbf{F}$$

主成分分析と因子分析は何が違うのか？

因子分析



主成分分析



$$X = AF + U$$

より表現能力の高いモデル
理論は複雑

違いは誤差項 U

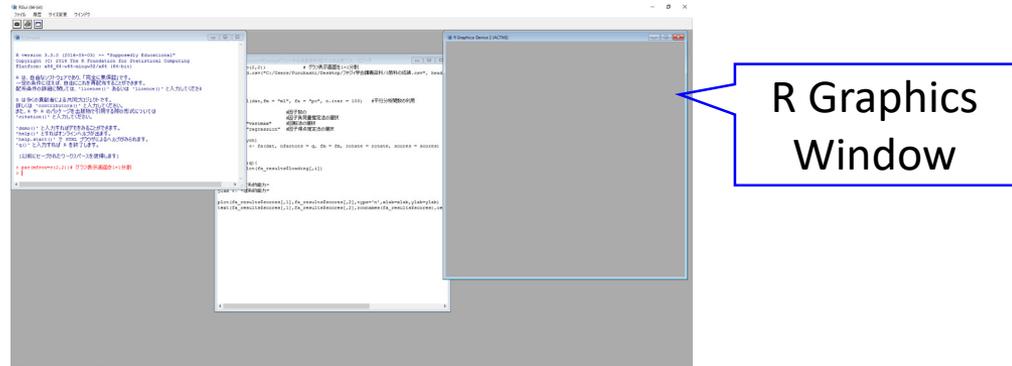
$$X = B^{-1}F$$

表現能力は限定的
理論は簡単

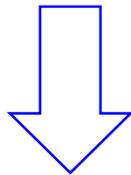
データ x_i を生成せしめた
因子のモデルを同定する.

主成分得点 f_i の分散を最大にする軸を同定する.

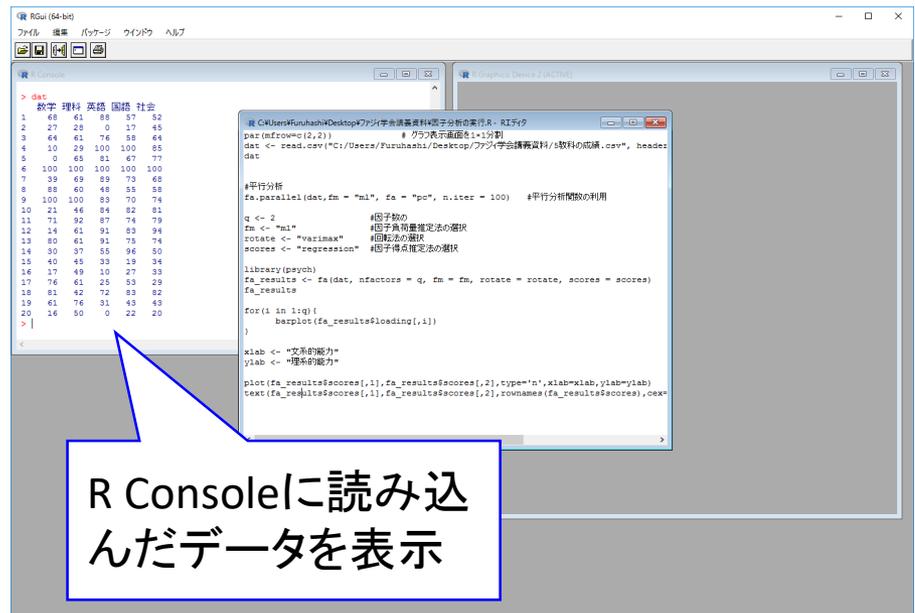
- (5) 先頭行のスク립トにカーソルを置き左クリック
- (6) Ctrl + Rにより先頭行を実行



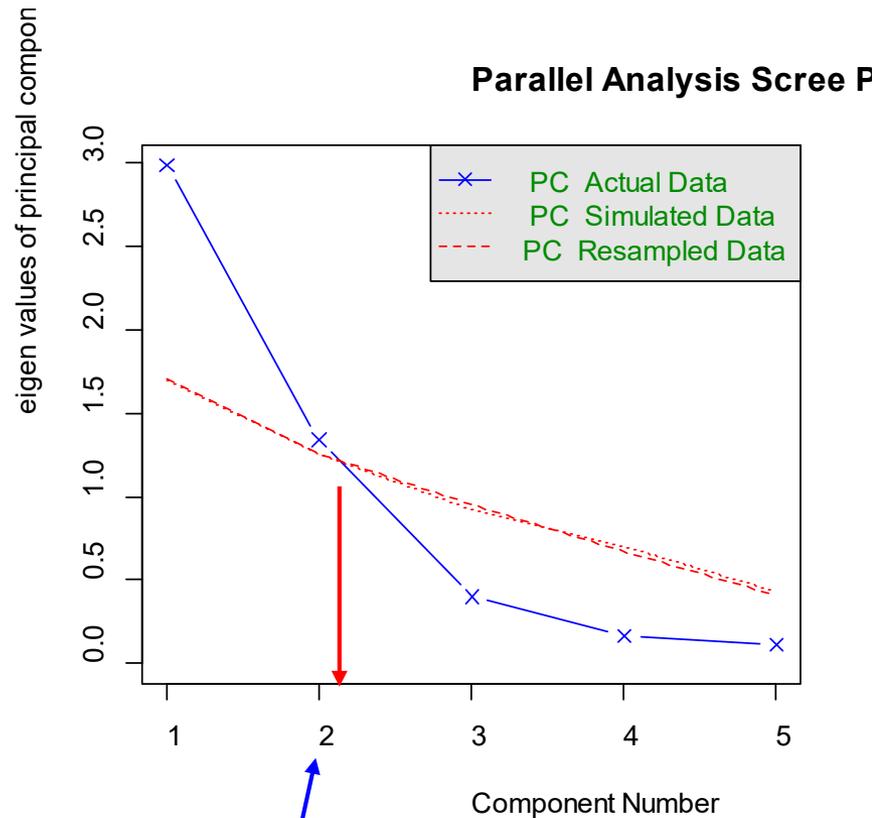
- (7) 第2行目のスク립トにカーソルを置き左クリック → Ctrl+Rにより実行
- (8) Ctrl + Rにより第3行目を実行



以降, Ctrl + Rを押す度にスク
リプトが1行ずつ実行される.



(9) #平行分析 `fa.parallel(dat, fm = "ml", fa = "pc", n.iter = 100)` を実行

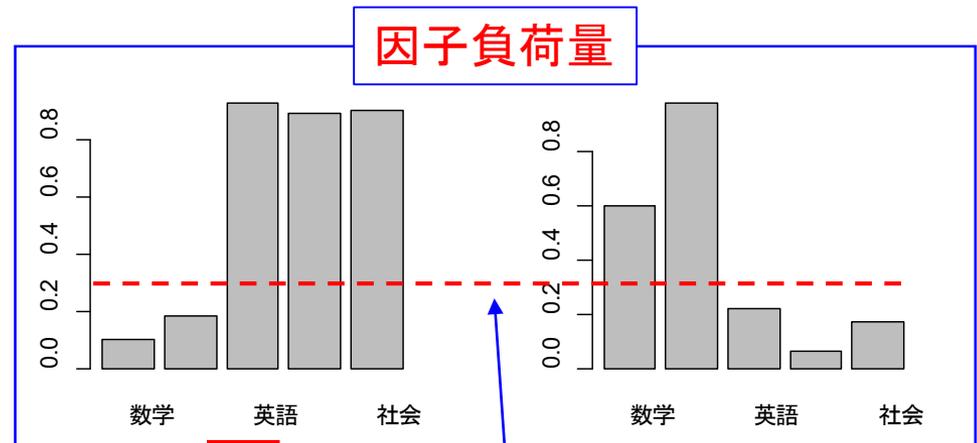


交点より左にある因子数として2を採用.

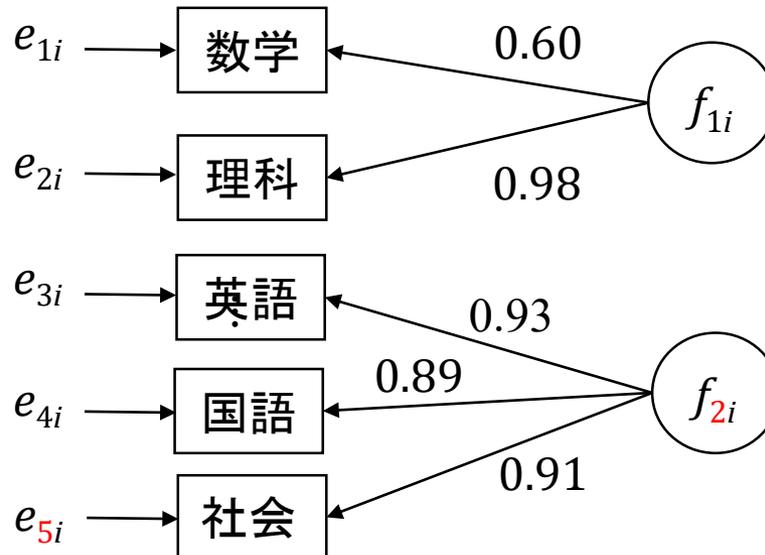
(10) Ctrl + Aにより全スクリプトを選択 → Ctrl + Rにより全スクリプトを実行

```
for(i in 1:q){  
  barplot(fa_results$loading[,i])  
}
```

により因子負荷量を棒グラフにより表示



得られた因子モデル



0.3より小さい因子負荷量を無視すると、左図の因子モデルが得られる。

scores <- "regression"

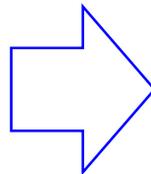
plot(fa_results\$scores[,1],fa_...)

text(fa_results\$scores[,1],fa_...)

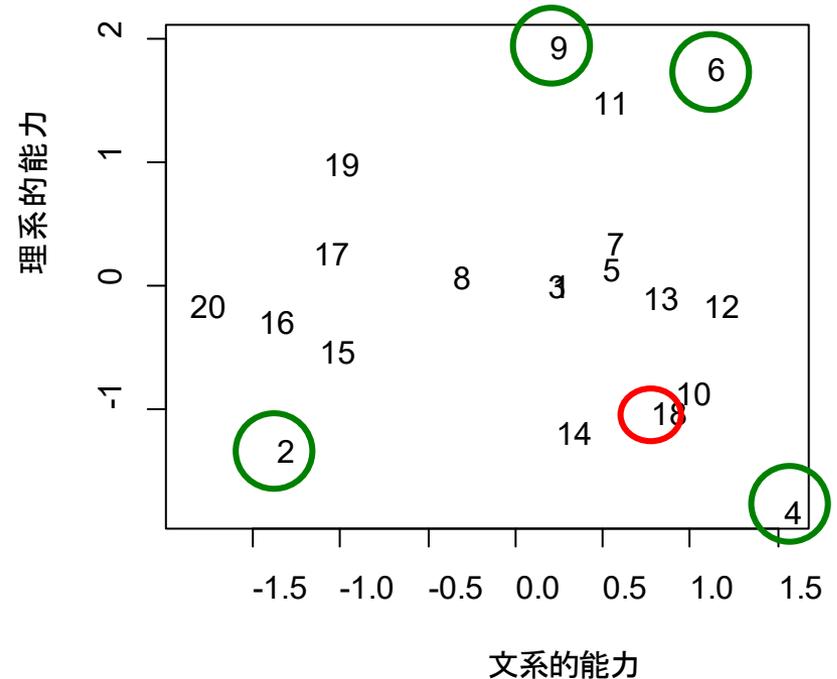
デフォルトの
因子得点推定法

により因子得点を表示

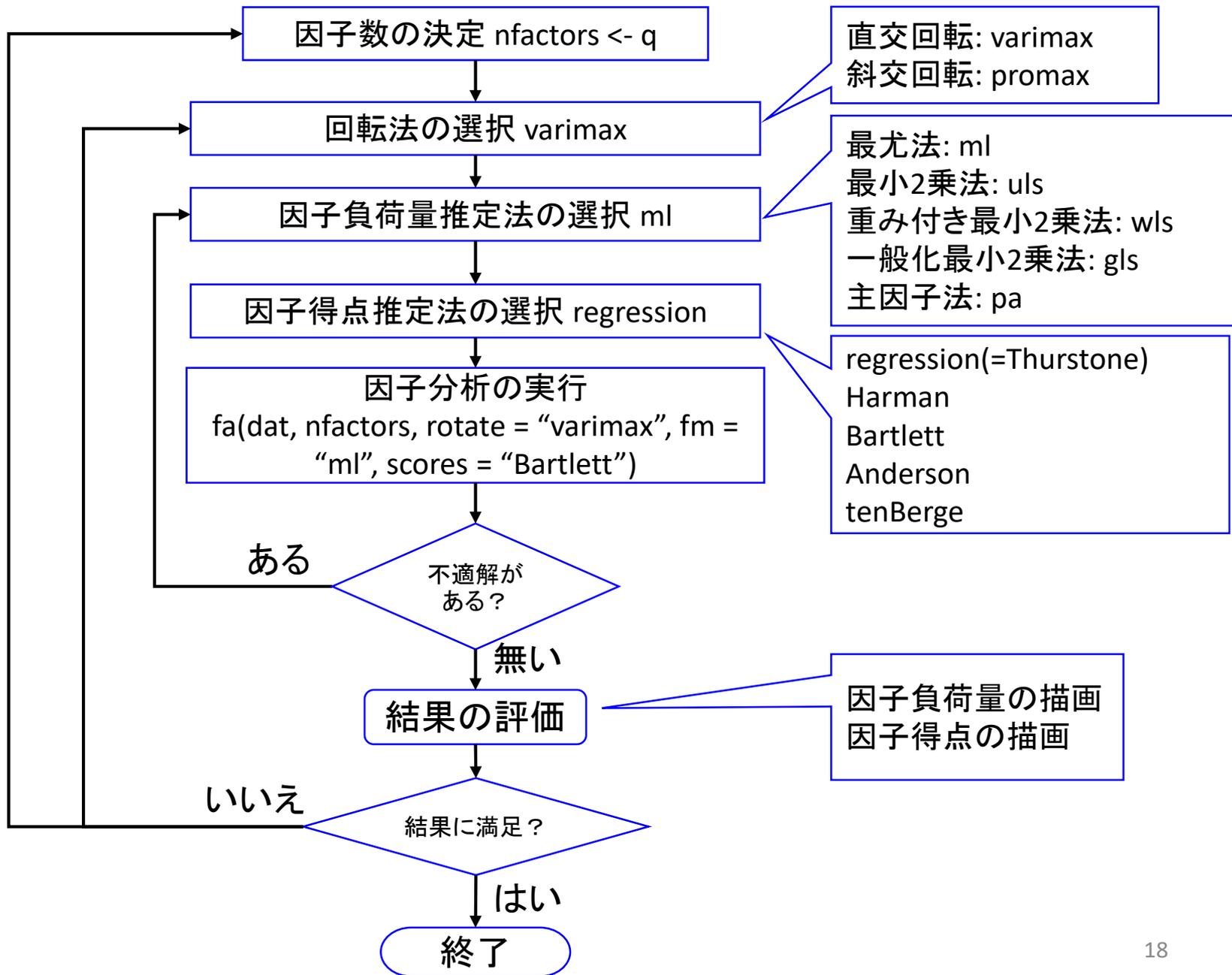
No.	数学	理科	英語	国語	社会
1	68	61	88	57	52
2	27	28	0	17	45
3	64	61	76	58	64
4	10	29	100	100	85
5	0	65	81	67	77
6	100	100	100	100	100
7	39	69	89	73	68
8	88	60	48	55	58
9	100	100	83	70	74
10	21	46	84	82	81
11	71	92	87	74	79
12	14	61	91	83	94
13	80	61	91	75	74
14	30	37	55	96	50
15	40	45	33	19	34
16	17	49	10	27	33
17	76	61	25	53	29
18	81	42	72	83	82
19	61	76	31	43	43
20	16	50	0	22	20



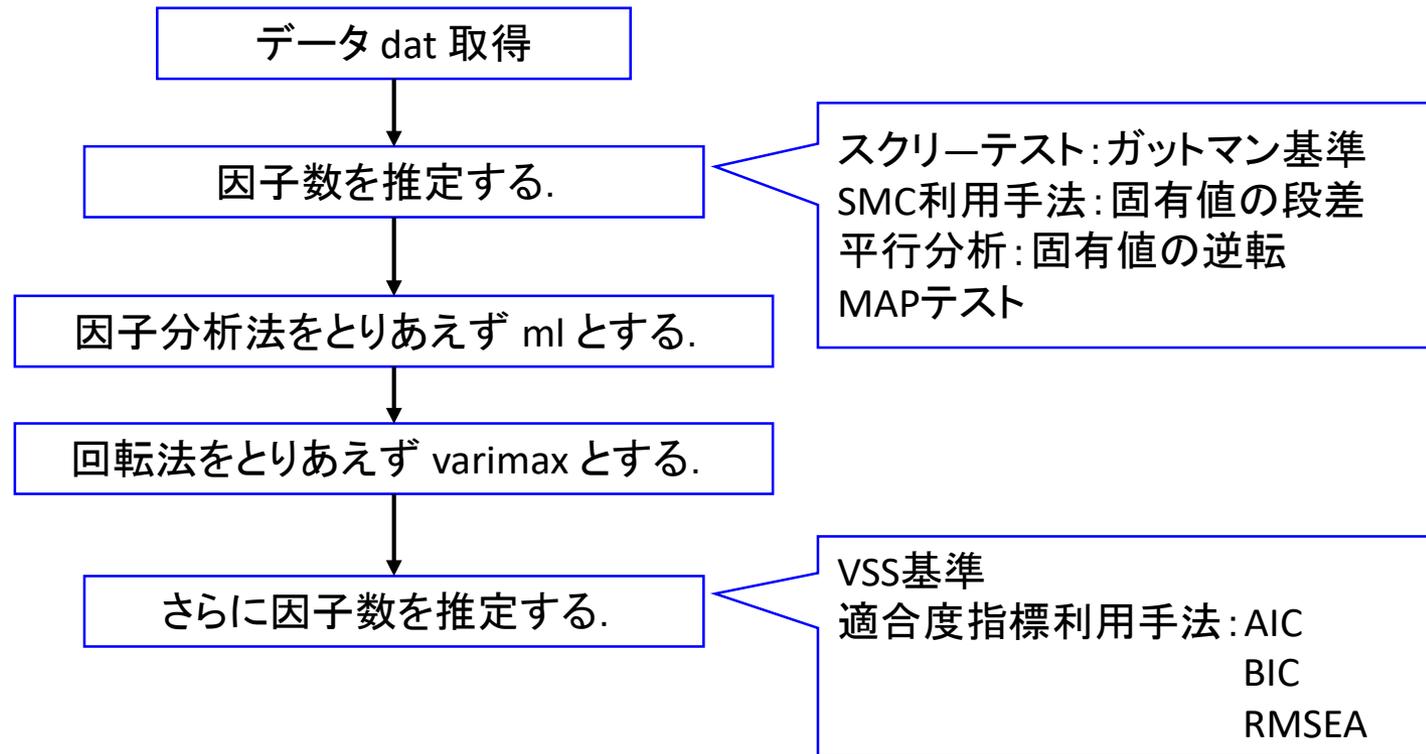
20人の生徒を文系/理系的能力因子
によりプロット



8. 因子分析の実行フロー



6. 因子数の決定法のフロー



同一のデータでも手法によって最適な因子数は異なる。適切な因子数を決定できる絶対的な方法はない。様々な方法を適用し、先行研究や実質科学的な観点からよく吟味することで因子数の決定を行いましょう。

7. 因子数の決定法を試してみよう

- (1) 「因子数の決定. R」を開く.
- (2) スクリーンテスト以下の3行を実行

データから得られた相関行列 S の固有値が急激に小さくなる一つ前の固有値の数を因子数とする.

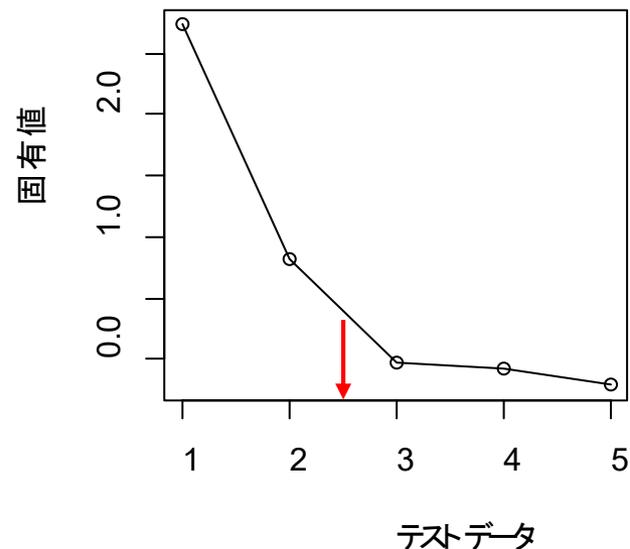
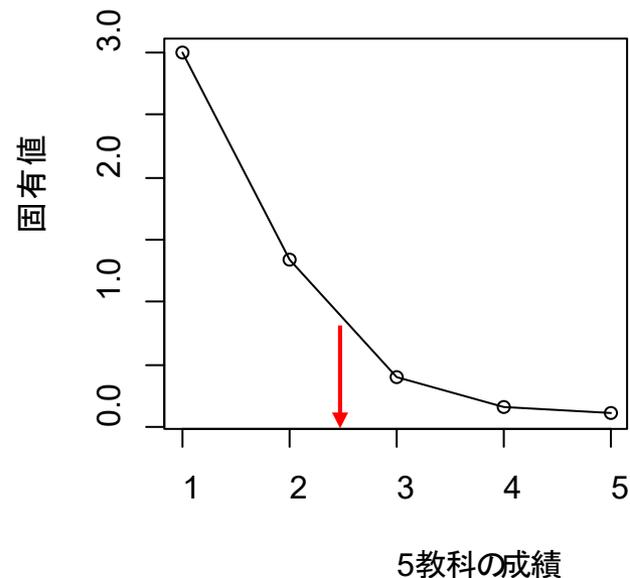
Guttman基準: 1以上の固有値の数を因子数とする.

- (3) SMC利用手法以下の5行を実行

0に近い固有値の中で、直前との差が大きい固有値の一つ前の固有値の数を因子数とする.

SMC: S^{-1} の対角要素を s_{jj} とすると $1 - 1/s_{jj}$ を対角要素とする行列.

$q/p \rightarrow 0$ のとき, $SMC = \Psi$ (誤差項) となる性質を持つ. 従って, $S - SMC$ は S よりも AA^T に近い値を持つ. Rのpsychライブラリには `smc()` 関数が用意されている.



(4) 平行分析の行を実行

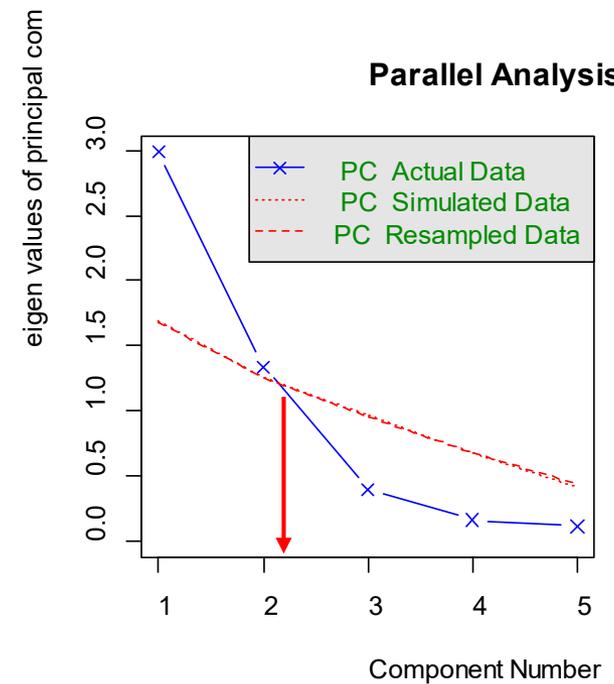
```
library(psych)
fa.parallel(dat, fm = "ml", fa = "pc", n.iter = 100)
```

により平行分析を実施.

X の各要素を乱数で置き換えて得られた X^* の相関行列 $S = X^*X^{*T}$ の固有値を求めて大きい順に λ_k ($k = 1, 2, \dots, p$)とする. $n.iter$ 回繰り返して, 各回の λ_k の平均値を求める.

S の固有値と比較して, 大きさが逆転する前の固有値までの個数を因子数とする.

乱数データを因子分析しても適切な解は得られない. 乱数データから得られる固有値より実データから得られる固有値が小さくなることは, その固有値により得られる結果は適切ではないと判断する.



(5) 当研究室の方法

因子数を決めて、因子分析を実施し、因子負荷量に意味が無い判断される因子が出たひとつ前の因子数を採用する。

因子数1

	ML1
数学	0.226
理科	0.375
英語	0.959
国語	0.881
社会	0.919

因子数2

	ML2	ML1
数学	0.103	0.603
理科	0.184	0.980
英語	0.929	0.221
国語	0.894	
社会	0.905	0.172

因子数3

	ML3	ML2	ML1
数学		0.616	
理科	0.188	0.974	-0.105
英語	0.892	0.227	
国語	0.920		0.374
社会	0.951	0.142	-0.181

複数の変数に関する因子を同定したい場合には、ひとつの因子にのみ関与する因子が出た時点で因子数が多すぎると判断する。

9.4 因子負荷量推定法の使い分け方

(1) デフォルトは ml

mlでは不適解に陥ることがある.

不適解: 共通性 $h_i^2 = 1 - d_i^2 > 1$ となった解
Heywood caseとも呼ばれる.

(2) $uls \rightarrow wls \rightarrow gls \rightarrow pa$ と試していく.

不適解に陥ることなく, 解が求まることがある.

特に pa は繰り返し手法であるため, 初期値設定によっては何らかの解が出る. その場合, 最適解ではなく, 途中で止まるだけの可能性があるが, 因子構造のヒントが得られるかも知れない.

9. 因子負荷量推定の考え方

$$\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} = \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 \end{pmatrix} + \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}$$

$S \approx \Sigma$

データから決まる

AA^t

決められない

ψ

決められない



基準を基に A, Ψ を決定する

残差平方和を最小化 簡単のため因子数を2とすると

$$\mathbf{S} - \mathbf{A}\mathbf{A}^t - \boldsymbol{\Psi} = \begin{pmatrix} 1 - a_1^2 - d_1^2 & r - a_1 a_2 \\ r - a_1 a_2 & 1 - a_2^2 - d_2^2 \end{pmatrix}$$

各項の2乗和をとると

$$(1 - a_1^2 - d_1^2)^2 + (r - a_1 a_2)^2 + (r - a_1 a_2)^2 + (1 - a_2^2 - d_2^2)^2$$

$$= \text{tr}(\mathbf{S} - \mathbf{A}\mathbf{A}^t - \boldsymbol{\Psi})^t (\mathbf{S} - \mathbf{A}\mathbf{A}^t - \boldsymbol{\Psi})$$

$$= \text{tr}(\mathbf{S} - \mathbf{A}\mathbf{A}^t - \boldsymbol{\Psi})^2 \quad \rightarrow \text{最小2乗法}$$

$$F(\mathbf{A}, \boldsymbol{\Psi}) = \text{tr}(\mathbf{S} - \mathbf{A}_1\mathbf{A}_1^t - \boldsymbol{\Psi})^2$$

$$\frac{\partial F(\mathbf{A}, \boldsymbol{\Psi})}{\partial \mathbf{A}} = 0$$

$$\frac{\partial F(\mathbf{A}, \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}} = 0$$

を求め

$$F = \text{tr}(\mathbf{S} - \mathbf{A}\mathbf{A}^t - \boldsymbol{\Psi})^2 \quad \boldsymbol{\Psi} = \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}$$

$$\frac{dF}{d\mathbf{A}} = \begin{pmatrix} \frac{d}{da_1} \{(1 - a_1^2 - d_1^2)^2 + (r - a_1 a_2)^2 + (r - a_1 a_2)^2 + (1 - a_2^2 - d_2^2)^2\} \\ \frac{d}{da_2} \{(1 - a_1^2 - d_1^2)^2 + (r - a_1 a_2)^2 + (r - a_1 a_2)^2 + (1 - a_2^2 - d_2^2)^2\} \end{pmatrix}$$

$$= -4 \begin{pmatrix} (1 - a_1^2 - d_1^2)a_1 + (r - a_1 a_2)a_2 \\ (r - a_1 a_2)a_1 + (1 - a_2^2 - d_2^2)a_2 \end{pmatrix}$$

$$= -4 \begin{pmatrix} 1 - a_1^2 - d_1^2 & r - a_1 a_2 \\ r - a_1 a_2 & 1 - a_2^2 - d_2^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

$$= -4(\mathbf{S} - \mathbf{A}\mathbf{A}^t - \boldsymbol{\Psi})\mathbf{A} = \mathbf{0}$$

$$\frac{dF}{d\boldsymbol{\Psi}} = \begin{pmatrix} \frac{dF}{d\psi_{11}} & \frac{dF}{d\psi_{12}} \\ \frac{dF}{d\psi_{21}} & \frac{dF}{d\psi_{22}} \end{pmatrix}$$

$$= -2 \begin{pmatrix} (1 - a_1^2 - d_1^2) & 0 \\ 0 & (1 - a_2^2 - d_2^2) \end{pmatrix}$$

$$= \mathbf{0}$$

$$\boldsymbol{\Psi} = \text{diag}(\mathbf{S} - \mathbf{A}\mathbf{A}^t)$$

$$(S - AA^t - \Psi)A = 0 \quad \Rightarrow \quad (S - \Psi)A = AA^tA$$

\Rightarrow $S - \Psi$ の固有値 $\theta_1, \theta_2, \dots, \theta_p$
 固有ベクトル $\omega_1, \omega_2, \dots, \omega_p$

AA^t は $S - \Psi$ の
 大きい方から q
 個の固有値か
 らなる。

\Rightarrow

$$\Omega_1 = (\omega_1 \omega_2 \cdots \omega_q)$$

$$\Theta_1 = \begin{pmatrix} \theta_1 & 0 & \cdots & 0 \\ 0 & \theta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta_q \end{pmatrix}$$

因子数の数 q
 だけ固有値, 固
 有ベクトルを取
 り出す。

\Rightarrow

$$\Omega_1 \Theta_1 \Omega_1^t A = AA^t A$$

$$\Omega_1 \Theta_1 \Omega_1^t = AA^t$$

$$\Omega_1 \Theta_1^{\frac{1}{2}} \Theta_1^{\frac{t}{2}} \Omega_1^t = AA^t$$

$$A = \Omega_1 \Theta_1^{\frac{1}{2}}$$

(1) $S - \Psi$ の初期値を S とする.

(2) $S - \Psi$ の固有値 $\theta_1, \theta_2, \dots, \theta_p$, 固有ベクトル $\omega_1, \omega_2, \dots, \omega_p$ を求める.

(3) 大きい方から q 個の固有値と, 対応する固有ベクトルを取り出して

$$\left. \begin{array}{l} \Omega_1 = (\omega_1 \omega_2 \cdots \omega_q) \\ \Theta_1 = \begin{pmatrix} \theta_1 & 0 & \cdots & 0 \\ 0 & \theta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta_q \end{pmatrix} \end{array} \right\} A = \Omega_1 \Theta_1^{\frac{1}{2}}$$

(4) AA^t とその対角要素(共通性)の総和を求めて, 更新前の総和との差の絶対値が閾値以下となれば終了. そうでなければ(2)へ戻る.

$$\text{diag}(S - \Psi) = \text{diag}(AA^t) \quad \leftarrow \quad \Psi = \text{diag}(S - AA^t)$$

「pa_source.R」のスキプトのコア部分

```
while (err > min.err) {  
  
  eigens <- eigen(r.mat)  
  
  loadings <- eigens$vector[, 1:nfactors] %*%  
    diag(sqrt(eigens$values[1:nfactors]))  
  
  model <- loadings %*% t(loadings)  
  
  new <- diag(model)  
  comm1 <- sum(new)  
  diag(r.mat) <- new  
  err <- abs(comm - comm1)  
  comm <- comm1  
  
  i <- i + 1  
}
```

誤差errがmin.err以下になるまで
while(){*}の*を反復実行

S - Ψの固有値・固有ベクトルの計算

$A = \Omega_1 \Theta_1^{\frac{1}{2}}$ の計算

model = $AA^t = (S - \Psi)$ の計算

diag(AA^t): 共通性
共通性の総和の計算
S - Ψの対角要素の更新
共通性の総和の差分の絶対値を計算
次の反復回のためにcommの更新

反復回数のカウントアップ

9.2 最尤法

「ml_source.R」のスク립ト

p 次元のデータ \mathbf{x}_i が平均 μ , 相関行列 Σ の正規母集団から無作為抽出されたとする. \mathbf{x}_i の確率密度関数は次式で与えられる.

$$f(\mathbf{x}_i) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} \mathbf{x}_i^t \Sigma^{-1} \mathbf{x}_i \right\}$$

n 個のデータ $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ が無作為抽出されたとする. $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ の同時確率密度関数は

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = f(\mathbf{x}_1) f(\mathbf{x}_2) \cdots f(\mathbf{x}_n)$$

$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \mathbf{x}_i^t \Sigma^{-1} \mathbf{x}_i \right\}$$

↓ 変形

$$= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left\{ -\frac{1}{2} \text{ntr}(\mathbf{S} \Sigma^{-1}) \right\}$$

因子分析は $\Sigma = AA^T + \Psi$ とするモデル. A と Ψ を求める指標を, 以下の確率密度関数とする. サンプルが得られれば, 相関行列 S が計算できる. この S の下で, サンプルデータが出現する確率を最大にする A と Ψ が最も尤もらしいとする考え方.

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp \left\{ -\frac{1}{2} \text{ntr}(\mathbf{S}\Sigma^{-1}) \right\}$$

$$\Sigma = AA^t + \Psi$$

このままでは扱いにくいので, 両辺の対数を取り, A と Ψ を含まない項を消去すると

$$\ln f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = -\frac{n}{2} \ln |\Sigma| - \frac{1}{2} \text{ntr}(\mathbf{S}\Sigma^{-1})$$

となる. 以下の変形をして,

$$\begin{aligned} F(\mathbf{A}, \Psi) &= -\frac{2}{n} \ln f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) - \ln |\mathbf{S}| \\ &= -\ln |\mathbf{S}\Sigma^{-1}| + \text{tr}(\mathbf{S}\Sigma^{-1}) \end{aligned}$$

これを最小とする A と Ψ を求める.

$$\Sigma = AA^t + \Psi$$

$F(\mathbf{A}, \Psi)$ の最小化問題



$$\left. \begin{aligned} \frac{\partial F(\mathbf{A}, \Psi)}{\partial \mathbf{A}} &= 0 \\ \frac{\partial F(\mathbf{A}, \Psi)}{\partial \Psi} &= 0 \end{aligned} \right\}$$

を求めて問題を整理する

$$\begin{aligned} -\frac{\partial}{\partial \mathbf{A}} \ln |\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}| &= \frac{-1}{|\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}|} \frac{\partial}{\partial \mathbf{A}} |\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}| \\ &= \frac{-1}{|\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}|} \frac{\partial}{\partial \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \left| \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix}^{-1} \right| \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \left| \mathbf{S} \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix}^{-1} \right| &= \frac{\partial}{\partial \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} |\mathbf{S}| \left| \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix}^{-1} \right| = \frac{\partial}{\partial \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} |\mathbf{S}| \left| \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix}^{-1} \right| \\ &= \frac{\partial}{\partial \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \frac{|\mathbf{S}|}{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2} = \frac{-|\mathbf{S}|}{((a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2)^2} \begin{pmatrix} (a_2^2 + d_2^2)2a_1 - a_2^2 2a_1 \\ (a_1^2 + d_1^2)2a_2 - a_1^2 2a_2 \end{pmatrix} \\ &= \frac{-2|\mathbf{S}|}{|\mathbf{A}\mathbf{A}^t + \Psi|^2} \begin{pmatrix} a_2^2 + d_2^2 & -a_1 a_2 \\ -a_1 a_2 & a_1^2 + d_1^2 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = -2|\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}|(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}\mathbf{A} \end{aligned}$$

$$\begin{aligned} -\frac{\partial}{\partial \mathbf{A}} \ln |\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}| &= \frac{1}{|\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}|} 2|\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}|(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}\mathbf{A} \\ &= 2(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}\mathbf{A} \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{A}} \text{tr}(\mathbf{S}(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1}) &= \frac{\partial}{\partial \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \text{tr} \left(\mathbf{S} \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix}^{-1} \right) \\
&= \frac{\partial}{\partial \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \text{tr} \left(\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \begin{pmatrix} a_2^2 + d_2^2 & -a_1 a_2 \\ -a_1 a_2 & a_1^2 + d_1^2 \end{pmatrix} \right) \\
&= \frac{\partial}{\partial \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \frac{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2}{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2} \\
&= \frac{\partial}{\partial \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}} \frac{a_2^2 + d_2^2 - r a_1 a_2 + a_1^2 + d_1^2 - r a_1 a_2}{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2} \\
&= \frac{1}{((a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2)^2} \left((2a_1 - 2ra_2)((a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2) - (a_2^2 + d_2^2 - ra_1 a_2 + a_1^2 + d_1^2 - ra_1 a_2)(2a_1(a_2^2 + d_2^2) - 2a_1 a_2^2) \right) \\
&= -2 \frac{\begin{pmatrix} a_2^2 + d_2^2 & -a_1 a_2 \\ -a_1 a_2 & a_1^2 + d_1^2 \end{pmatrix}}{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2} \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \frac{\begin{pmatrix} a_2^2 + d_2^2 & -a_1 a_2 \\ -a_1 a_2 & a_1^2 + d_1^2 \end{pmatrix}}{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \\
&= -2(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1} \mathbf{S}(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1} \mathbf{A}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial F}{\partial \mathbf{A}} &= -\frac{\partial}{\partial \mathbf{A}} \ln \left| \mathbf{S}(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1} \right| + \frac{\partial}{\partial \mathbf{A}} \text{tr} \left(\mathbf{S}(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1} \right) \\
&= 2(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1} \mathbf{A} - 2(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1} \mathbf{S}(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1} \mathbf{A} \\
&= 2(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1} (\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi} - \mathbf{S})(\mathbf{A}\mathbf{A}^t + \boldsymbol{\Psi})^{-1} \mathbf{A}
\end{aligned}$$

$$-\frac{\partial}{\partial \Psi} \ln |\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}| = \frac{-1}{|\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}|} \frac{\partial}{\partial \Psi} |\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}|$$

$$= \frac{-1}{|\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}|} \frac{\partial}{\partial \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}} \left| \mathbf{S} \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix}^{-1} \right|$$

$$\begin{aligned} \frac{\partial}{\partial \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}} \left| \mathbf{S} \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix}^{-1} \right| &= \frac{\partial}{\partial \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}} |\mathbf{S}| \left| \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix}^{-1} \right| = \frac{\partial}{\partial \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}} |\mathbf{S}| \left| \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix} \right|^{-1} \\ &= \frac{\partial}{\partial \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}} \frac{|\mathbf{S}|}{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2} = \frac{-|\mathbf{S}|}{((a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2)^2} \begin{pmatrix} a_2^2 + d_2^2 & 0 \\ 0 & a_1^2 + d_1^2 \end{pmatrix} \\ &= \frac{-|\mathbf{S}|}{|\mathbf{A}\mathbf{A}^t + \Psi|} \text{diag}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1} = -|\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}| \text{diag}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1} \end{aligned}$$

$$\begin{aligned} -\frac{\partial}{\partial \Psi} \ln |\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}| &= \frac{1}{|\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}|} |\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}| \text{diag}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1} \\ &= \text{diag}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1} \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \Psi} \text{tr}(\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1}) &= \frac{\partial}{\partial \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}} \text{tr} \left(\mathbf{S} \begin{pmatrix} a_1^2 + d_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 + d_2^2 \end{pmatrix}^{-1} \right) \\
&= \frac{\partial}{\partial \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}} \text{tr} \left(\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \begin{pmatrix} a_2^2 + d_2^2 & -a_1 a_2 \\ -a_1 a_2 & a_1^2 + d_1^2 \end{pmatrix} \right) \\
&= \frac{\partial}{\partial \begin{pmatrix} d_1^2 & 0 \\ 0 & d_2^2 \end{pmatrix}} \frac{a_2^2 + d_2^2 - r a_1 a_2 + a_1^2 + d_1^2 - r a_1 a_2}{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2} \\
&= \begin{pmatrix} \frac{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2 - (a_2^2 + d_2^2 - r a_1 a_2 + a_1^2 + d_1^2 - r a_1 a_2)(a_2^2 + d_2^2)}{((a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2)^2} & 0 \\ 0 & \frac{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2 - (a_2^2 + d_2^2 - r a_1 a_2 + a_1^2 + d_1^2 - r a_1 a_2)(a_1^2 + d_1^2)}{((a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2)^2} \end{pmatrix} \\
&= -\text{diag} \left(\frac{\begin{pmatrix} a_2^2 + d_2^2 & -a_1 a_2 \\ -a_1 a_2 & a_1^2 + d_1^2 \end{pmatrix}}{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2} \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \frac{\begin{pmatrix} a_2^2 + d_2^2 & -a_1 a_2 \\ -a_1 a_2 & a_1^2 + d_1^2 \end{pmatrix}}{(a_1^2 + d_1^2)(a_2^2 + d_2^2) - a_1^2 a_2^2} \right) \\
&= -\text{diag}((\mathbf{A}\mathbf{A}^t + \Psi)^{-1} \mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial F}{\partial \Psi} &= -\frac{\partial}{\partial \Psi} \ln \left| \mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1} \right| + \frac{\partial}{\partial \Psi} \text{tr} \left(\mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1} \right) \\
&= \text{diag}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1} - \text{diag} \left((\mathbf{A}\mathbf{A}^t + \Psi)^{-1} \mathbf{S}(\mathbf{A}\mathbf{A}^t + \Psi)^{-1} \right) \\
&= \text{diag} \left((\mathbf{A}\mathbf{A}^t + \Psi)^{-1} (\mathbf{A}\mathbf{A}^t + \Psi - \mathbf{S})(\mathbf{A}\mathbf{A}^t + \Psi)^{-1} \right) = 0
\end{aligned}$$

$$\left. \begin{aligned} \frac{\partial}{\partial \mathbf{A}} \ln |\mathbf{S}\boldsymbol{\Sigma}^{-1}| &= 2\boldsymbol{\Sigma}^{-1}\mathbf{A} \\ \frac{\partial}{\partial \mathbf{A}} \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) &= 2\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{A} \end{aligned} \right\} \frac{\partial F(\mathbf{A}, \boldsymbol{\Psi})}{\partial \mathbf{A}} = -2\boldsymbol{\Sigma}^{-1}\mathbf{A} + 2\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}\mathbf{A} = 0$$

$$\left. \begin{aligned} \frac{\partial}{\partial \boldsymbol{\Psi}} \ln |\mathbf{S}\boldsymbol{\Sigma}^{-1}| &= \text{diag}\boldsymbol{\Sigma}^{-1} \\ \frac{\partial}{\partial \boldsymbol{\Psi}} \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) &= -\text{diag}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1}) \end{aligned} \right\} \frac{\partial F(\mathbf{A}, \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}} = \text{diag}(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma} - \mathbf{S})\boldsymbol{\Sigma}^{-1}) = 0$$

→

$$F(\Psi) = \sum_{j=m+1}^p (-\log\theta_j + \theta_j) + q - p$$

$$\frac{\partial F}{\partial \Psi} = \Psi^{-\frac{1}{2}} \text{diag}(AA^t + \Psi - S) \Psi^{-\frac{1}{2}}$$

準ニュートン法により F を最小とする Ψ, A を求める.

ただし, $\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}$ の固有値を $\theta_1, \theta_2, \dots, \theta_p$, 固有ベクトル $\omega_1, \omega_2, \dots, \omega_p$ とすると

$$\Rightarrow \Omega_1 = (\omega_1 \omega_2 \cdots \omega_q) \Rightarrow A = \Psi^{-1} \Omega_1 (\Theta_1 - I_q)^{\frac{1}{2}}$$

因子数の数 q だけ固有値, 固有ベクトルを取り出す.

$$\Theta_1 = \begin{pmatrix} \theta_1 & 0 & \cdots & 0 \\ 0 & \theta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta_q \end{pmatrix}$$

fa()関数の中のスクリプト

「ml_source.R」のスクリプト

$$F(\Psi) = \sum_{j=m+1}^p (-\log\theta_j + \theta_j) + q - p \quad \text{を計算する関数}$$

```
FAfn <- function(Psi, S, q) {  
  sc <- diag(1/sqrt(Psi))
```

$$sc = \text{diag}(\Psi^{-\frac{1}{2}})$$

```
  Sstar <- sc %*% S %*% sc
```

$$S^* = \Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}$$

```
  E <- eigen(Sstar, symmetric = TRUE,  
            only.values = TRUE)
```

S^* の固有値計算
 S^* は対称行列
固有値のみを求める。

```
  e <- E$values[-(1:q)]
```

E に格納された固有値のうち、大きいほうから q 個を除いた固有値 θ_j を e に格納

```
  e <- sum(log(e) - e) - m + nrow(S)
```

$$F = \sum_{j=m+1}^p (-\log\theta_j + \theta_j) + q - p$$

```
  -e
```

```
}
```

fa()関数の中のスクリプト

$$\frac{\partial F}{\partial \Psi} = \Psi^{-\frac{1}{2}} \text{diag}(AA^t + \Psi - S) \Psi^{-\frac{1}{2}} \text{ を計算する関数}$$

```
FAgr <- function(Psi, S, q) {
  sc <- diag(1/sqrt(Psi))
```

$$sc = \text{diag}(\Psi^{-\frac{1}{2}})$$

```
Sstar <- sc %**% S %**% sc
```

$$S^* = \Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}$$

```
E <- eigen(Sstar, symmetric = TRUE)
```

S^* の固有値計算, S^* は対称行列
固有値, 固有ベクトルを求める。

```
Omega1 <- E$vector[, 1:q, drop = FALSE]
```

Ω_1 : S^* の大きいほうから q 個の固有値に対応する
固有ベクトル

```
Theta1 <- pmax(E$values[1:q] - 1, 0)
```

θ_1 : S^* の大きいほうから q 個の固有値. 固有値が負
の値の場合は0を入れる。

```
load <- Omega1 %**% diag(sqrt(Theta1), q)
```

$$A = \Psi^{\frac{1}{2}} \Omega_1 (\Theta_1 - I_q)^{\frac{1}{2}}$$

```
A <- diag(sqrt(Psi)) %**% load
```

```
g <- A %**% t(A) + diag(Psi) - S
```

$$\frac{\partial F}{\partial \Psi} = \Psi^{-\frac{1}{2}} \text{diag}(AA^t + \Psi - S) \Psi^{-\frac{1}{2}}$$

```
diag(g)/Psi^2
```

```
}
```

fa()関数の中のスクリプト

準ニュートン法を計算する関数 optim()

```
"fit" <- function(S, q, fm) {
```

```
  S.smc <- smc(S)
```

```
  start <- diag(S) - S.smc
```

```
  res <- optim(start, FAFn, FAgr, method = "L-BFGS-B", ...
```

smc()関数の利用. $q/p \rightarrow 0$ のとき,
SMC = Ψ (誤差項)となる性質を持つ.
従って, $S - SMC$ を Ψ の初期値とする.

準ニュートン法のLimited memory (メモリ節約型(計算機に向けた))BFGS法. B (Bounded)は Ψ の探索範囲に上下限を与えられる方法.

$$\frac{\partial F}{\partial \Psi} = \Psi^{-\frac{1}{2}} \text{diag}(AA^t + \Psi - S) \Psi^{-\frac{1}{2}}$$

$$F(\Psi) = \sum_{j=m+1}^p (-\log \theta_j + \theta_j) + q - p$$

9.3 その他の因子負荷量推定法

最小2乗法 $F(\mathbf{A}, \mathbf{\Psi}) = \text{tr}(\mathbf{S} - \mathbf{A}_1\mathbf{A}_1^t - \mathbf{\Psi})^2$ ⇨ 残差平方和最小

「uls_source.R」のSCRIPT

重み付き最小2乗法 $F(\mathbf{A}, \mathbf{\Psi}) = \text{tr}(\mathbf{Sd}(\mathbf{S} - \mathbf{A}\mathbf{A}^t - \mathbf{\Psi})^2\mathbf{Sd})$

「wls_source.R」のSCRIPT

ただし $\mathbf{Sd} = \begin{pmatrix} 1/s_{11} & 0 & \cdots & 0 \\ 0 & 1/s_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/s_{pp} \end{pmatrix}$

⇨ 重み付き残差平方和最小

s_{jj} は \mathbf{S}^{-1} の対角要素

共通性の低い項目をより重みづける効果

一般化最小2乗法 $F(\mathbf{A}, \mathbf{\Psi}) = \text{tr}(\mathbf{S}^{-1}(\mathbf{S} - \mathbf{A}\mathbf{A}^t - \mathbf{\Psi})^2)$

「gls_source.R」のSCRIPT

⇨ 最小化

共通性の低い項目をより重みづける効果

10. 因子の回転

10.1 直交回転 varimax

$$S = AA^t + \Psi$$



$$\Psi = \text{diag}(S - AA^t)$$

直交行列 T を考える.

直交行列とは

$$T^t T = I$$

を満たす行列. T は回転行列.

$A' = AT^t$ としてベクトル A を回転させる.

$$\begin{aligned}\Psi &= \text{diag}(S - A'A'^t) \\ &= \text{diag}(S - AT^tTA^t) \\ &= \text{diag}(S - AA^t)\end{aligned}$$

Ψ は回転に対して不定

基準の決め方により, 各種回転法が提案されている.

p 変数で因子数 q 個のとき

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix}$$

Varimax基準

$$E = \sum_{j=1}^q \left\{ \frac{1}{p} \sum_{i=1}^p (b_{ij}^2 - \bar{b}^2_j)^2 \right\}$$

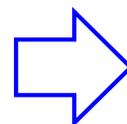
因子負荷量の2乗値の分散の和

$$\text{ただし, } \bar{a}^2_j = \frac{1}{p} \sum_{i=1}^p b_{ij}^2$$

因子負荷量の2乗値の平均値

Varimax回転の考え方

$$T = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \cos\theta & \dots & -\sin\theta & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \sin\theta & \dots & \cos\theta & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$



$$B = AT$$

Varimax基準を最大にする θ を求める.

この回転を全ての列の組合せ(${}_q C_2$ 通り)について実行する. 一通り実行し終えても, 全体として最大化されているとは限らないので, ${}_q C_2$ 通りの最大化を B が収束するまで反復する.

$$E = \sum_{j=1}^q \left\{ \frac{1}{p} \sum_{i=1}^p (b_{ij}^2 - \overline{b^2_j})^2 \right\} \quad \text{ただし, } \overline{b^2_j} = \frac{1}{p} \sum_{i=1}^p b_{ij}^2$$

E の両辺に p を掛けて平方を開いて整理すると

$$U = \sum_{j=1}^q \left\{ \frac{1}{p} \sum_{i=1}^p b_{ij}^4 - \frac{1}{p} \left(\sum_{i=1}^p b_{ij}^2 \right)^2 \right\}$$

この評価関数を, 変換行列 T の正規直交条件の制約の下で最大化する.

$$Q = U - 2 \sum_{j=1}^q \lambda_{jj} (\mathbf{t}_j^t \mathbf{t}_j - 1) - 2 \sum_{i \neq j}^q \lambda_{ij} \mathbf{t}_i^t \mathbf{t}_j$$

Q を, 変換行列 T の各成分で偏微分すると

$$\frac{\partial Q}{\partial \mathbf{t}_j} = \begin{pmatrix} \frac{\partial Q}{\partial t_{1j}} \\ \vdots \\ \frac{\partial Q}{\partial t_{qj}} \end{pmatrix} \quad \frac{\partial Q}{\partial t_{lj}} = 4 \sum_{i=1}^p b_{ij}^3 a_{il} - \frac{4}{p} \left(\sum_{i=1}^p b_{ij}^2 \right) \sum_{i=1}^p b_{ij}^3 a_{il} - 4\lambda_{lj} t_{lj} = 0$$

まとめると

$$A^t B^{(3)} - T\Lambda = 0 \quad B^{(3)} = \left(b_{ij}^3 - \frac{4}{p} b_{ij} \sum_{k=1}^p b_{kj}^2 \right)$$

よって

$$B = AT$$

$$A^t B^{(3)} = T\Lambda$$

を満足する B を求めればよいが、直接解くことは困難。そこで、両式から Λ を消去して

$$B = AT$$

$$T = A^t B^{(3)} \left(B^{(3)t} A A^t B^{(3)} \right)^{-\frac{1}{2}}$$

T に適当な初期値 (単位行列) を与えて B, T を求めることを T が収束するまで繰り返す。

「varimax_source.R」のスク립ト

```
T <- diag(q)          # Tを単位行列とする.
d <- 0
for(i in 1L:1000L) {
  B <- A %*% T        #  $B = AT$ 
  TT <- t(A) %*% (B^3 - B %*% diag(drop(rep(1, p) %*% B^2))/p)
  #  $A^t B^{(3)} = A^t \left( b_{ij}^3 - \frac{4}{p} b_{ij} \sum_{k=1}^p b_{kj}^2 \right)$ 
  inv_TT <- solve(t(TT)%*%TT) #  $\left( B^{(3)t} A A^t B^{(3)} \right)^{-1}$ 
  inv_eig <- eigen(inv_TT)
  T <- TT %*% inv_eig$vectors %*% diag(sqrt(inv_eig$values)) %*% t(inv_eig$vectors)
  #  $T = A^t B^{(3)} \left( B^{(3)t} A A^t B^{(3)} \right)^{-\frac{1}{2}}$ 
  sTT <- La.svd(TT)
  dpast <- d
  d <- sum(sTT$d)
  if(d < dpast * (1 + eps)) break
  n_rep <- n_rep + 1
}
B <- A %*% T
```

収束条件の判定

10.2 斜交回転 promax

promax = varimax回転+斜交プロクラステス回転

最尤法などにより得られた因子負荷量ベクトルに対してまず varimax回転を施す.

$$B = AT$$

B の各要素を4乗することで, 仮説行列 C を作成

promax基準 $E = (C - BA)^2$ 最小とする回転行列 A を求める.

$$A = QD$$

$$Q = (B^t B)^{-1} B^t C$$

$$D = \{\text{diag}(Q^t Q)^{-1}\}^{\frac{1}{2}}$$

回転の結果得られる因子負荷量行列を $B_p (= B\Lambda)$ とすると,

$$\begin{aligned} S &= AA^t + \Psi \\ &= BB^t + \Psi \\ &= B_p \Lambda^{-1} \Lambda^{-t} B_p + \Psi \\ &= B_p \Phi B_p + \Psi \end{aligned}$$

となる. すなわち

$$\Phi = \Lambda^{-1} \Lambda^{-t} = FF^t \neq I$$

であり, $FF^t = I$ (因子負荷量は無相関) の仮定を崩すことを許容している.

「promax_source.R」のスク립ト

#斜交プロクラステス回転

```
r <- 4
```

```
C <- PA * abs(PA)^(r-1)
```

```
Q <- solve(t(PA) %*% PA) %*% t(PA) %*% C
```

```
D <- diag(sqrt(diag(solve(t(Q) %*% Q))))
```

```
Lambda <- Q %*% D
```

```
Pro1 <- PA %*% Lambda
```

```
#仮説行列Cの計算
```

```
#  $Q = (B^t B)^{-1} B^t C$ 
```

```
#  $D = \{\text{diag}(Q^t Q)^{-1}\}^{\frac{1}{2}}$ 
```

```
#  $\Lambda = QD$ 
```

```
# 回転後の因子負荷量 $BA$ の計算
```

10.3 rによる計算

「因子分析の実行(回転).R」ファイル

```
> rotate <- "none"
```

Loadings:

	ML2	ML1
数学		0.611
理科		0.997
英語	0.857	0.421
国語	0.858	0.258
社会	0.845	0.367

```
> rotate <- "varimax"
```

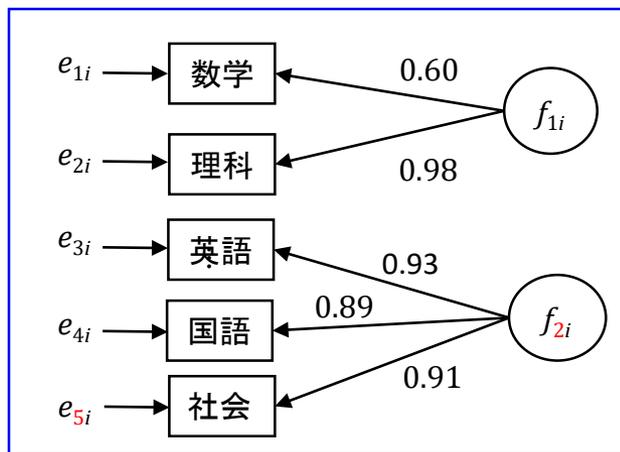
Loadings:

	ML2	ML1
数学	0.103	0.603
理科	0.184	0.980
英語	0.929	0.221
国語	0.894	
社会	0.905	0.172

```
> rotate <- "promax"
```

Loadings:

	ML2	ML1
数学		0.612
理科		0.993
英語	0.929	
国語	0.921	
社会	0.913	



11. 因子得点の推定

$$S = AA^t - \Psi$$

データから決まる

因子負荷量推定法(ml, uls, ...)により決定される.



$$X = AF + U$$

データから決まる

因子負荷量推定法(ml, uls, ...)により決定される.

サンプルサイズが n で
既知数が $p \times n + p \times q$ 個
未知数が $p \times n + q \times n$ 個
変数の数 $p <$ データ数 n なので
連立方程式は不定

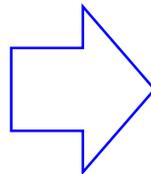
11.1 直交回転の場合

```
rotate <- "varimax"  
scores <- "regression"  
plot(fa_results$scores[,1],fa_...)  
text(fa_results$scores[,1],fa_...)
```

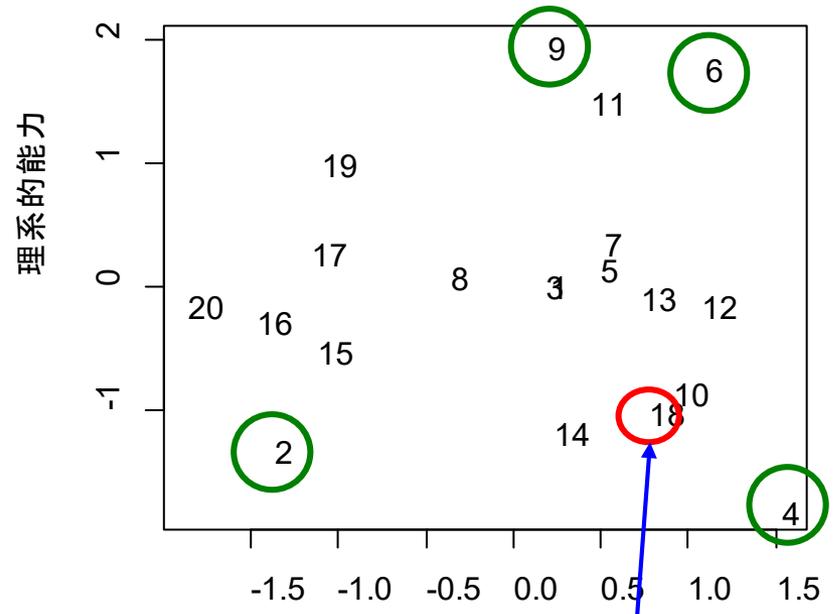
デフォルトの
因子得点推定法

により因子得点を表示

No.	数学	理科	英語	国語	社会
1	68	61	88	57	52
2	27	28	0	17	45
3	64	61	76	58	64
4	10	29	100	100	85
5	0	65	81	67	77
6	100	100	100	100	100
7	39	69	89	73	68
8	88	60	48	55	58
9	100	100	83	70	74
10	21	46	84	82	81
11	71	92	87	74	79
12	14	61	91	83	94
13	80	61	91	75	74
14	30	37	55	96	50
15	40	45	33	19	34
16	17	49	10	27	33
17	76	61	25	53	29
18	81	42	72	83	82
19	61	76	31	43	43
20	16	50	0	22	20



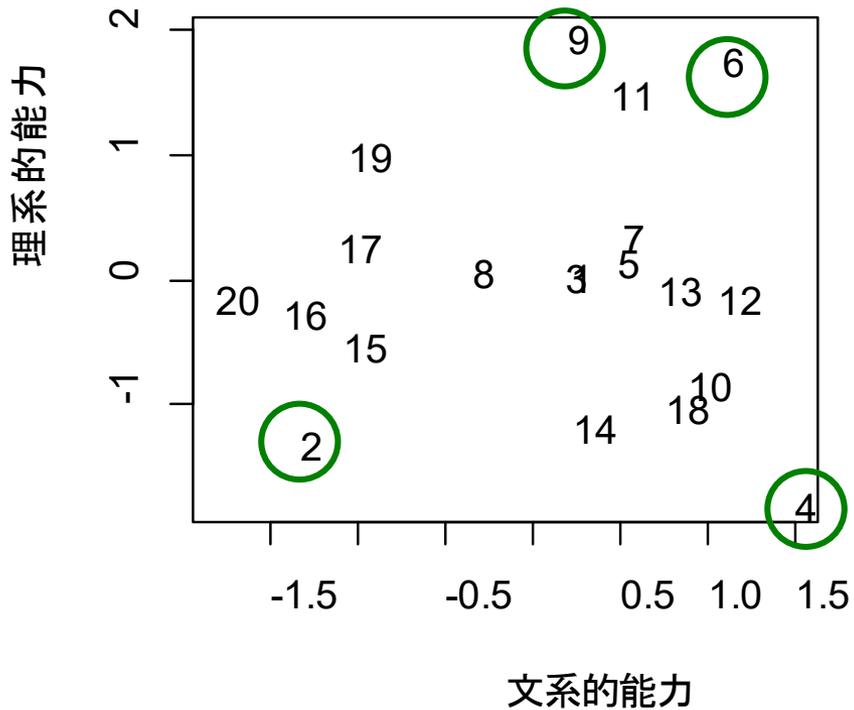
20人の生徒を文系/理系的能力因子によりプロット



Xの特徴とずれている

rotate <- "varimax"
 scores <- "Regression"

X と $A\hat{F}$ の差を
 最も小さくする
 \hat{F} を求める。

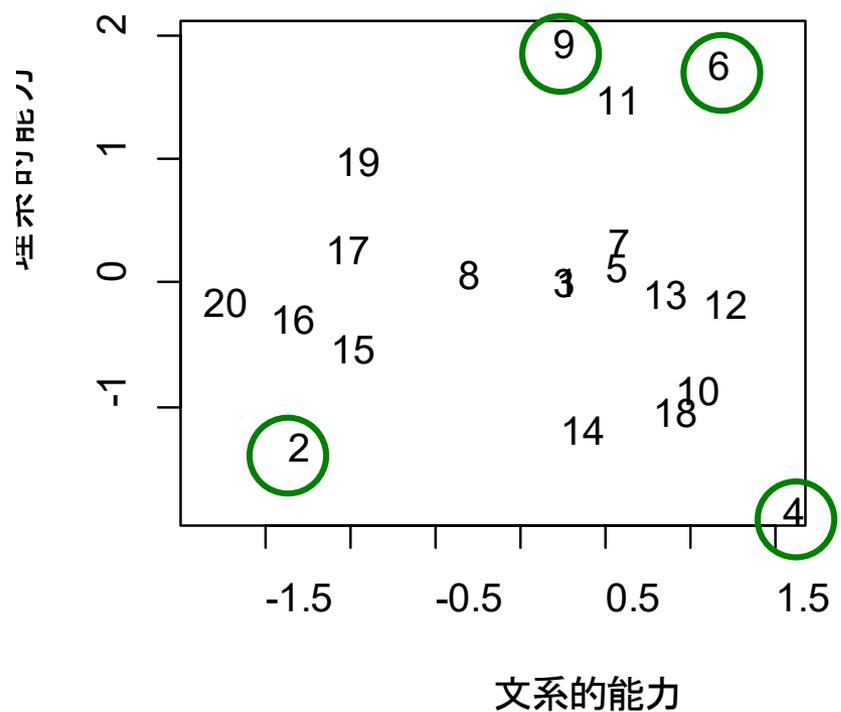


因子間相関

	ML2	ML1
ML2	1	0.0111
ML1	0.0111	1

rotate <- "varimax"
 scores <- "Anderson"

因子間を独立にする
 因子得点推定法



因子間相関

	ML2	ML1
ML2	1	0
ML1	0	1

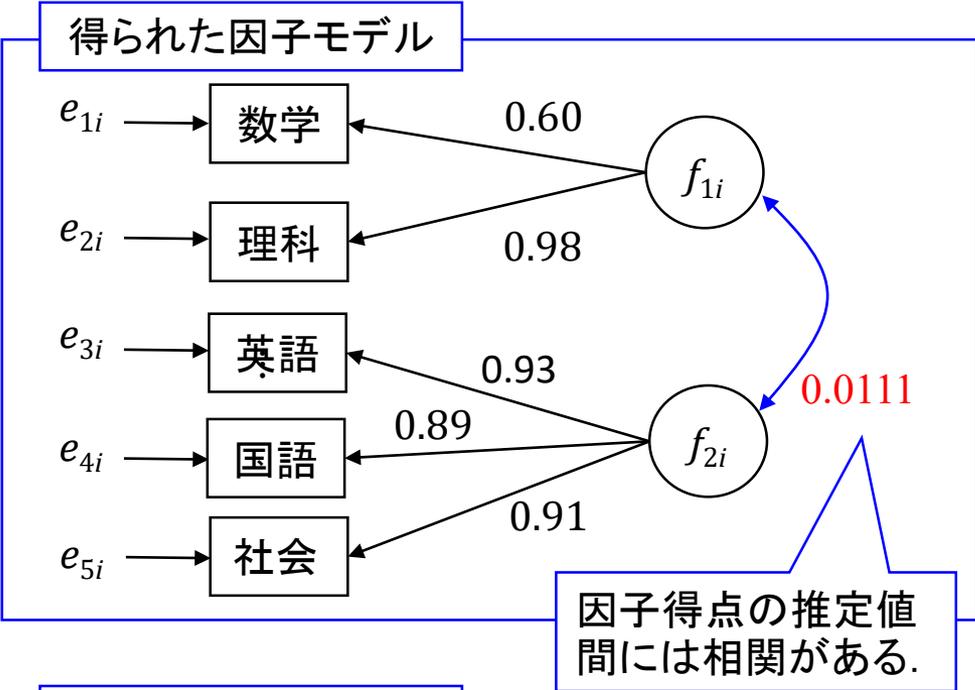
因子得点の相関
 値がゼロ

rotate = "varimax"
 scores = "regression"
 のとき

因子間相関

	ML2	ML1
ML2	1	0.0111
ML1	0.0111	1

因子間の相関係数



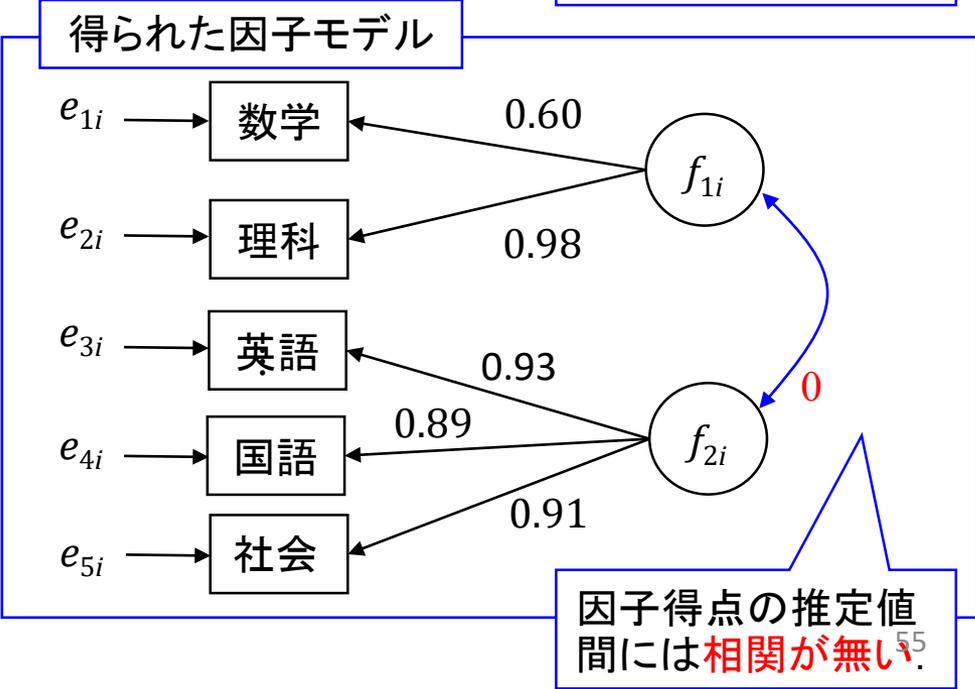
rotate = "varimax"
 scores = "Anderson"
 のとき

因子間を独立にする因子得点推定法

因子間相関

	ML2	ML1
ML2	1	0
ML1	0	1

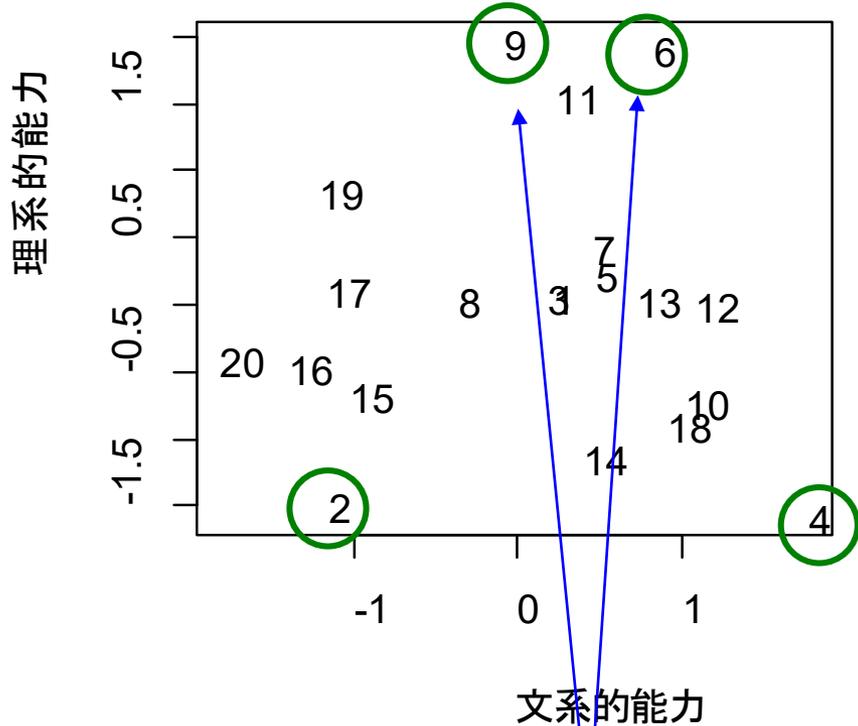
因子間の相関係数



11.2 斜交回転の場合

rotate <- "Promax"
scores <- "Anderson"

斜交回転の結果に
因子間を独立にする
因子得点推定法
を適用すると



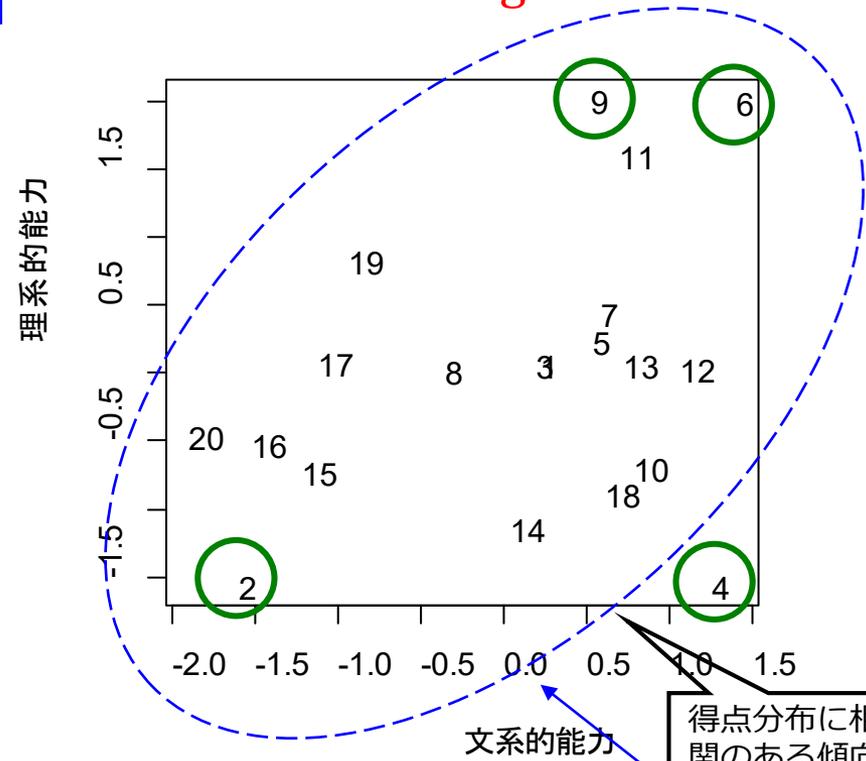
因子間相関

	ML2	ML1
ML2	1	0
ML1	0	1

斜交回転の結果に対
してAndersonを適用
すると、因子間相関
を0とするために、因
子得点がゆがむ。

rotate <- "Promax"
scores <- "tenBerge"

斜交回転の結果に因
子間の相関 ϕ を保存
する因子得点推定法
を適用すると



因子間相関

	ML2	ML1
ML2	1	0.3287
ML1	0.8287	1

得点分布に相
関のある傾向
が見える。

因子得点
間相関 ϕ が
保存され
ている。

Regression (Thurstone)

$$X = AF + U$$

因子得点の
真の値

$$\hat{F}^t = X^t w$$

因子得点の
推定値

因子得点の推定値を
 w を係数とする X の回
帰式により近似する。

$$\begin{aligned} E &= (F^{*t} - \hat{F}^t)^2 \\ &= (F^{*t} - X^t w)^2 \\ &= (F^{*t} - X^t w)^t (F^{*t} - X^t w) \\ &= F^* F^{*t} - F^* X^t w - w X F^{*t} + w^t X X^t w \end{aligned}$$

$$\frac{\partial E}{\partial w} = -2X F^{*t} + 2X X^t w = 0$$

因子得点の真の値と推定値の差の2乗
和を最小とする w を求める。

ここで $X = AF^* + U$ に右から F^{*t} をかけると

$$X F^{*t} = \underbrace{A F^* F^{*t}}_{I} + \underbrace{U F^{*t}}_{0} = A \text{より}$$

I

0



$$X X^t w = X F^{*t} = A$$

$$w = (X X^t)^{-1} A = S^{-1} A$$

$$\hat{F}^t = X^t w = X^t S^{-1} A$$

「regression_source.R」ファイル

fa()関数内のスクリプトの抜粋

A <- fa_results\$loadings #fa()関数出力の因子負荷量をAに格納

Xt <- scale(dat) #データを平均0, 分散1に規格化

S <- cor(dat) #データの相関行列 $S = XX^t$

w <- solve(S, A) # $w = S^{-1}A$

Ft <- Xt %*% w # $\hat{F}^t = X^t w$

Harman

$$X = AF + U$$

$$\begin{aligned} E &= (X - A\hat{F})^2 \\ &= (X - A\hat{F})^t (X - A\hat{F}) \\ &= X^t X - X^t A\hat{F} - \hat{F}^t A^t X + \hat{F}^t A^t A\hat{F} \end{aligned}$$

$$\frac{\partial E}{\partial \hat{F}} = -2A^t X + 2A^t A\hat{F} = 0$$

X と A \hat{F} の差を最も小さくする \hat{F} を求める。

$$\hat{F}^t = X^t A (A^t A)^{-1}$$

Bartlett

誤差項 Ψ の大きな項には、逆数の重みをつけることで \hat{F} への影響を減らしている。

$$\begin{aligned} E &= \left(\Psi^{-\frac{1}{2}} (X - A\hat{F}) \right)^2 \\ &= (X - A\hat{F})^t \Psi^{-1} (X - A\hat{F}) \\ &= X^t \Psi^{-1} X - X^t \Psi^{-1} A\hat{F} - \hat{F}^t A^t \Psi^{-1} X + \hat{F}^t A^t \Psi^{-1} A\hat{F} \end{aligned}$$

$$\frac{\partial E}{\partial \hat{F}} = -2A^t \Psi^{-1} X + 2A^t \Psi^{-1} A\hat{F} = 0$$

$$\hat{F}^t = X^t \Psi^{-1} A (A^t \Psi^{-1} A)^{-1}$$

Anderson and Rubin

因子間相関をゼロにする制約.
 Λ はラグランジュの未定乗数

$$\begin{aligned} E &= \left(\Psi^{-\frac{1}{2}}(X - A\hat{F}) \right)^2 + \Lambda(\hat{F}\hat{F}^t - I) \\ &= (X - A\hat{F})^t \Psi^{-1}(X - A\hat{F}) + \Lambda(\hat{F}\hat{F}^t - I) \\ &= X^t \Psi^{-1} X - X^t \Psi^{-1} A\hat{F} - \hat{F}^t A^t \Psi^{-1} X + \hat{F}^t A^t \Psi^{-1} A\hat{F} + \Lambda(\hat{F}\hat{F}^t - I) \end{aligned}$$

$$\frac{\partial E}{\partial \hat{F}} = -2A^t \Psi^{-1} X + 2A^t \Psi^{-1} A\hat{F} + 2\Lambda\hat{F} = 0$$

$$(A^t \Psi^{-1} A + \Lambda)\hat{F} = A^t \Psi^{-1} X$$

$$\hat{F} = (A^t \Psi^{-1} X (A^t \Psi^{-1} X)^t)^{-\frac{1}{2}} A^t \Psi^{-1} X$$

$$\hat{F}^t = X^t \Psi^{-1} A (A^t \Psi^{-1} S \Psi^{-1} A)^{-\frac{1}{2}}$$

fa()関数内のスクリプトの抜粋

A <- fa_results\$loadings

Xt <- scale(dat)

S <- cor(dat)

Phi = diag(fa_results\$uniqueness)

Phinv = solve(Phi)

#fa()関数出力の因子負荷量をAに格納

#データを平均0, 分散1に規格化

#データの相関行列

Ψ の項をPhiに格納

Ψ^{-1}

X <- t(A) %*% Phinv %*% S %*% Phinv %*% A

E <- eigen(X)

inv_sq_E <- E\$vectors %*% diag(1/sqrt(E\$values)) %*% t(E\$vectors)

$(A^t \Psi^{-1} S \Psi^{-1} A)^{-\frac{1}{2}}$ の計算

Ft <- Xt %*% Phinv %*% A %*% inv_sq_E

$$\hat{F}^t = X^t \Psi^{-1} A (A^t \Psi^{-1} S \Psi^{-1} A)^{-\frac{1}{2}}$$

参考文献

- [1] 豊田秀樹編著「Rで学ぶ最新データ解析: 因子分析入門」東京図書, 2012
- [2] 柳井, 繁樹, 前川, 市川「因子分析 - その理論と方法 -」朝倉書店, 1990
- [3] 芝祐順「因子分析法第2版」東京大学出版会, 1979

本稿の詳細な解説は

[多変量解析の基礎 IV \(因子分析\) - 理論とRによる演習 -](#)

と題してアマゾンよりkindle版として出版しています.