多変量解析

第14回　因子分析

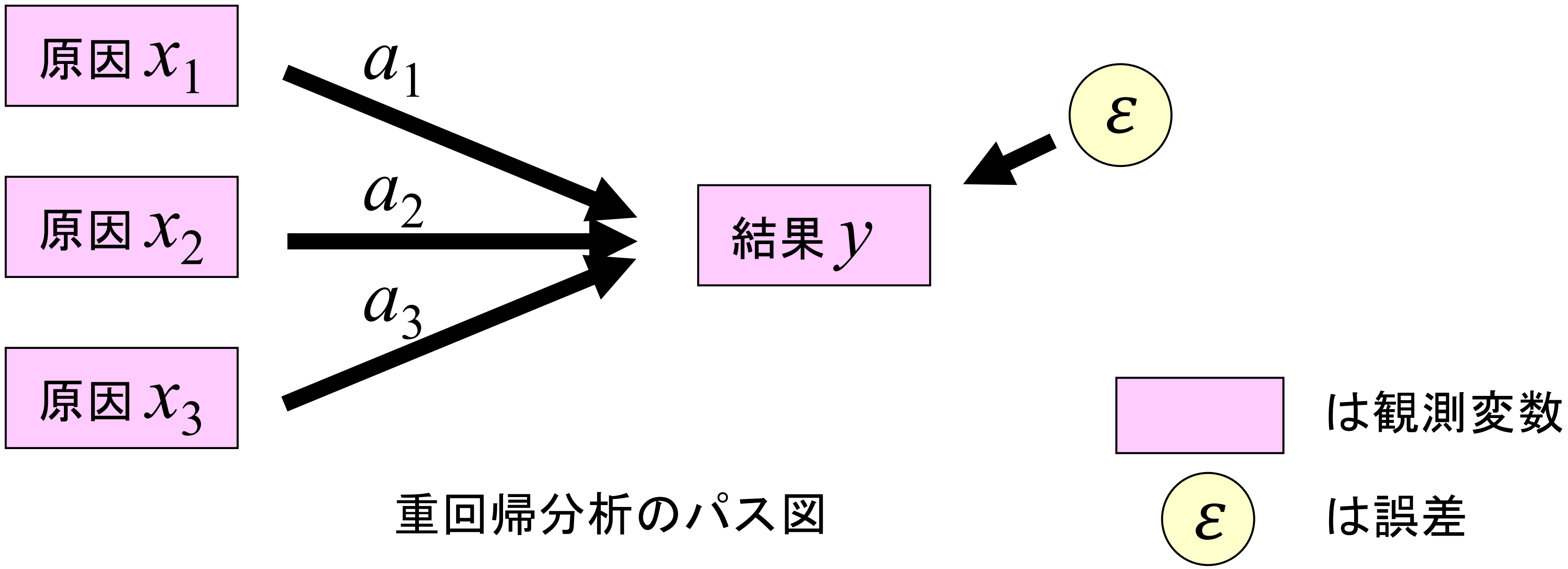
萩原•篠田
情報理工学部

重回帰分析

重回帰分析：複数の原因（説明変数）と結果（目的変数）を結ぶもの

目的変数 y （結果）と，それに影響を与えるいくつかの
説明変数 $x_1, x_2, \cdots x_p$ （原因）から

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \varepsilon$$



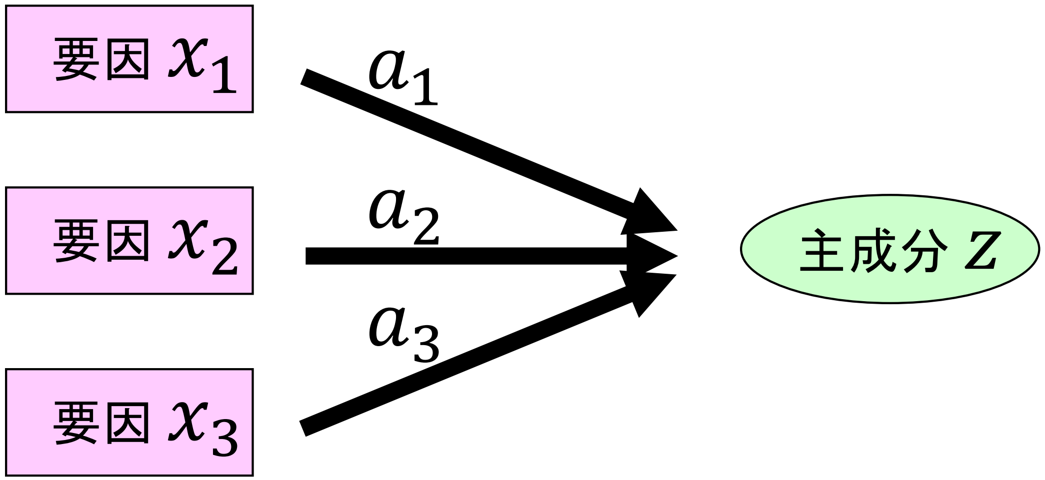
重回帰分析のパス図

主成分分析

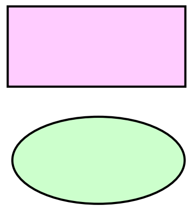
多くの要因（変数）($x_1, x_2, \cdots, x_p$)の値をできるだけ情報の損失なしに1個または互いに独立な少数の総合的指標($z_1, z_2, \cdots, z_m$)で表す。
つまり要因の総合化が主成分分析。

$$z = a_1x_1 + a_2x_2 + a_3x_3$$

主成分



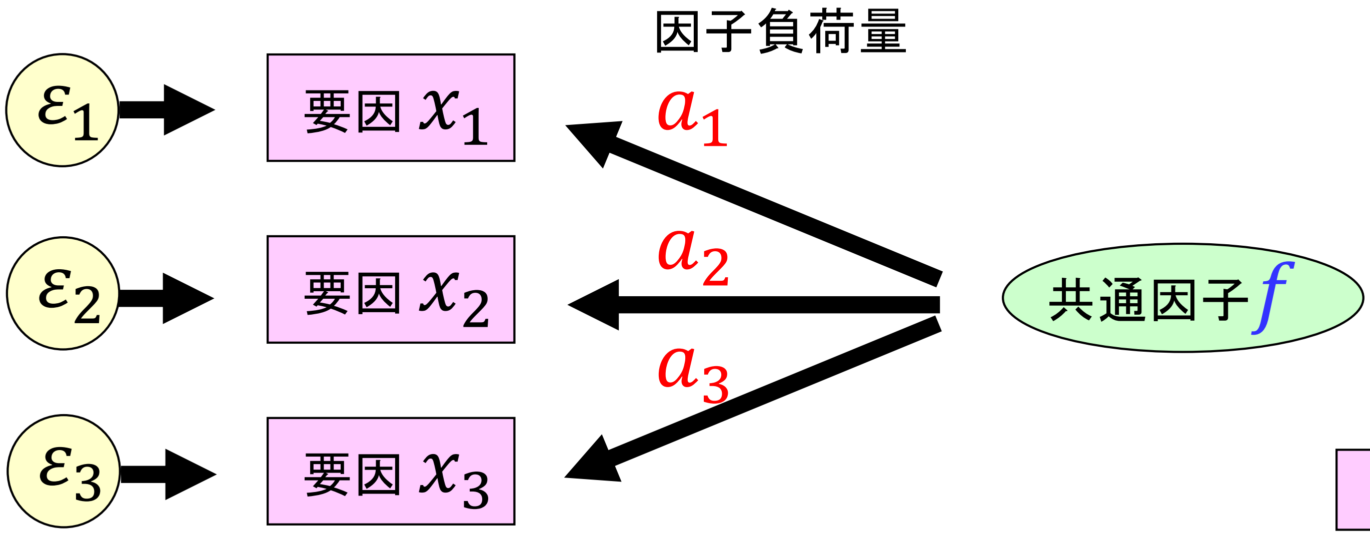
主成分分析のパス図



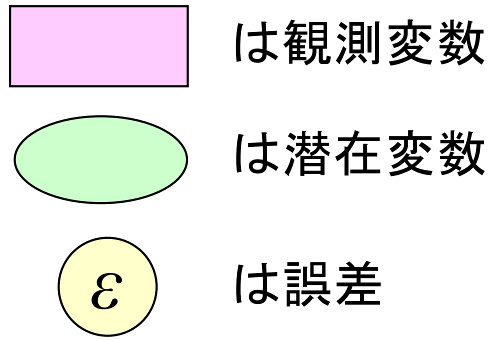
因子分析

実験や調査で得られたデータの構造がどのような構造になっているか
少数の特定な共通因子で説明する

因子分析；いくつかの要因の共通因子の抽出

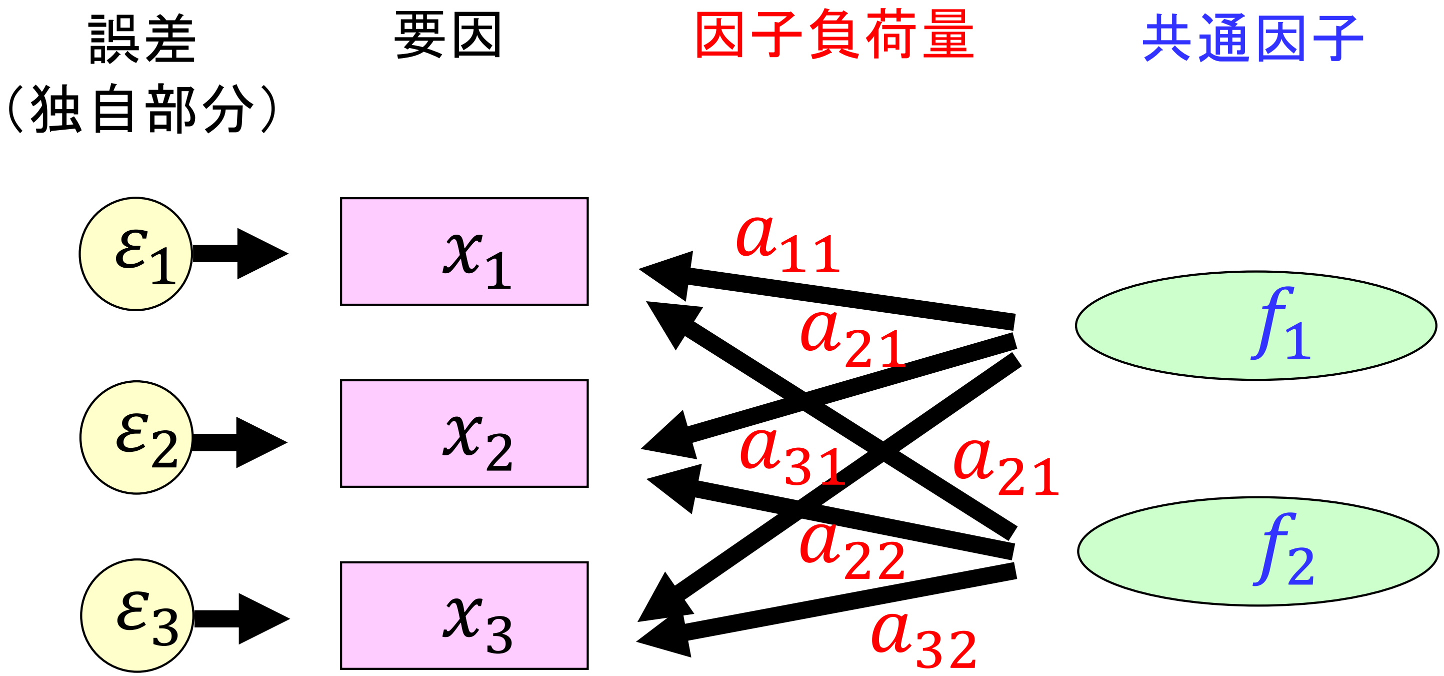


因子分析のパス図

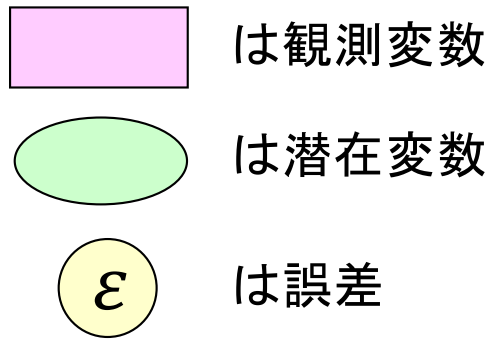


$$x_1 = a_1 f + \varepsilon_1$$
$$x_2 = a_2 f + \varepsilon_2$$
$$x_3 = a_3 f + \varepsilon_3$$

共通因子が2個の場合



因子分析のパス図



$$x_1 = a_{11}f_1 + a_{12}f_2 + \varepsilon_1$$
$$x_2 = a_{21}f_1 + a_{22}f_2 + \varepsilon_2$$
$$x_3 = a_{31}f_1 + a_{32}f_2 + \varepsilon_3$$

２店のコンビニの評価

要因 / 店名	品揃え ($x_{品}$)	雰囲気 ($x_{雰}$)	親近感 ($x_{親}$)	広々感 ($x_{広}$)
A店 (A)	56 ($x_{A品}$)	54 ($x_{A雰}$)	48 ($x_{A親}$)	46 ($x_{A広}$)
B店 (B)	44 ($x_{B品}$)	56 ($x_{B雰}$)	42 ($x_{B親}$)	44 ($x_{B広}$)

アンケート等データとして得る

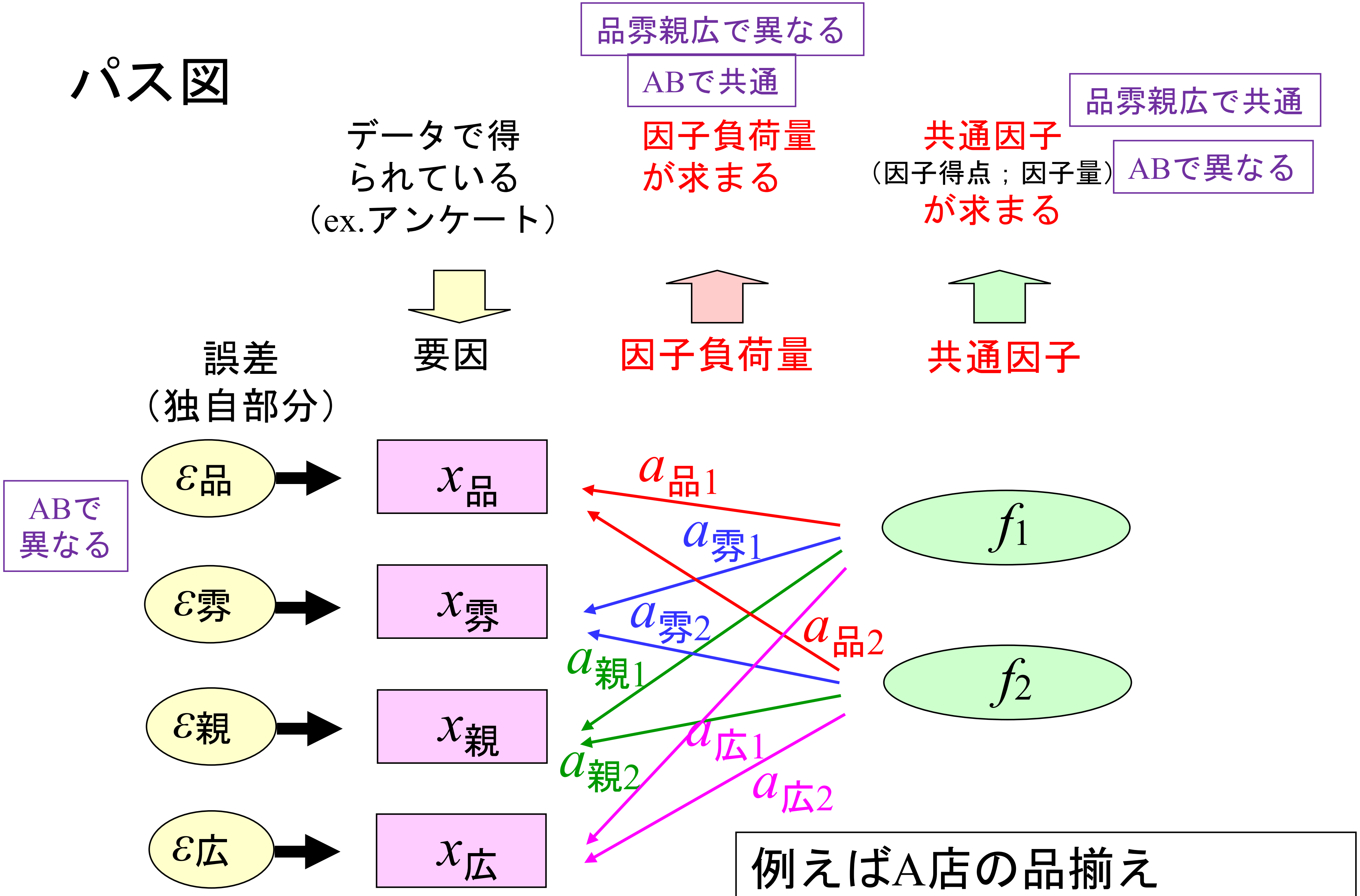
因子分析では、少数の<u>共通因子</u>があると仮定する

$\boxed{\text{実際の因子分析では仮定するのは因子数だけ}}$

今、共通因子が2個あると仮定する

共通因子　f_1、f_2

パス図



例えばA店の品揃え

$$x_{A品} = a_{品1} f_{A1} + a_{品2} f_{A2} + \varepsilon_{A品}$$

因子負荷量:因子の各変量(要因)への影響度

ABで共通　　品雰親広で異なる

すべての変量は共通因子1(f_1)、共通因子2(f_2)に関係する

変量	共通因子1(f_1)に関わる 因子負荷量 a_1	共通因子2(f_2)に関わる 因子負荷量 a_2
品揃え ($x_{品}$)	0.8 ($a_{品1}$)	0.2 ($a_{品2}$)
雰囲気 ($x_{雰}$)	0.5 ($a_{雰1}$)	0.6 ($a_{雰2}$)
親近感 ($x_{親}$)	0.6 ($a_{親1}$)	0.3 ($a_{親2}$)
広々感 ($x_{広}$)	0.5 ($a_{広1}$)	0.4 ($a_{広2}$)

例えば

$$x_{A品} = a_{品1} f_{A1} + a_{品2} f_{A2} + \varepsilon_{A品}$$
$$x_{A雰} = a_{雰1} f_{A1} + a_{雰2} f_{A2} + \varepsilon_{A雰}$$

$$x_{B品} = a_{品1} f_{B1} + a_{品2} f_{B2} + \varepsilon_{B品}$$
$$x_{B雰} = a_{雰1} f_{B1} + a_{雰2} f_{B2} + \varepsilon_{B雰}$$

ABで異なる　品雰親広で共通

2店のコンビニの共通因子の因子得点

店名	共通因子1の因子得点 (f_1)	共通因子2の因子得点 (f_2)
A店 (A)	60 (f_{A1})	40 (f_{A2})
B店 (B)	40 (f_{B1})	60 (f_{B2})

因子得点を保有していると仮定すると･･･

→ A店はf_1因子が大

→ B店はf_2因子が大

例えば

$$x_{A品} = a_{品1} f_{A1} + a_{品2} f_{A2} + \varepsilon_{A品}$$
$$x_{A雰} = a_{雰1} f_{A1} + a_{雰2} f_{A2} + \varepsilon_{A雰}$$

$$x_{B品} = a_{品1} f_{B1} + a_{品2} f_{B2} + \varepsilon_{B品}$$
$$x_{B雰} = a_{雰1} f_{B1} + a_{雰2} f_{B2} + \varepsilon_{B雰}$$

評価の決定

例えばA店の品揃え

$$x_{A品} = a_{品1} f_{A1} + a_{品2} f_{A2} + \varepsilon_{A品}$$

変量の評価点(要因)＝f_1因子負荷量×f_1因子得点
＋f_2因子負荷量×f_2因子得点

A店の品揃え評価点（要因）＝ 0.8 × 60 ＋ 0.2 × 40 ＝48＋8＝56

$$x_{A品} = a_{品1} \times f_{A1} + a_{品2} \times f_{A2}$$

A店の評価
品揃え＝56点
雰囲気＝54点
親近感＝48点
広々感＝46点

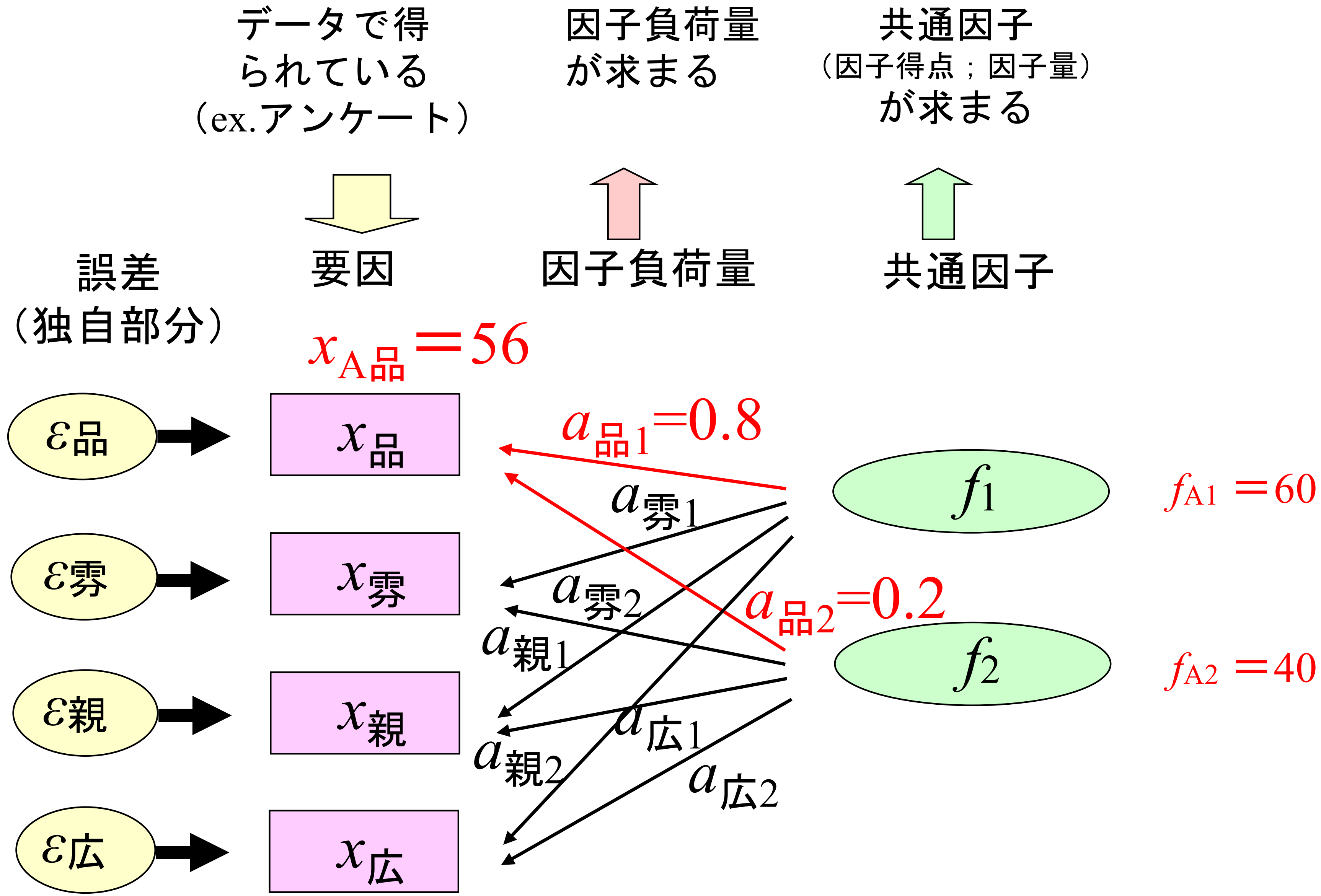
B店の評価
品揃え＝44点
雰囲気＝56点
親近感＝42点
広々感＝44点

店名	品揃え ($x_{品}$)	雰囲気 ($x_{雰}$)	親近感 ($x_{親}$)	広々感 ($x_{広}$)
A店 (A)	56 ($x_{A品}$)	54 ($x_{A雰}$)	48 ($x_{A親}$)	46 ($x_{A広}$)
B店 (B)	44 ($x_{B品}$)	56 ($x_{B雰}$)	42 ($x_{B親}$)	44 ($x_{B広}$)

一致

独自部分を0と仮定

パス図


データで得
られている
（ex.アンケート）
因子負荷量
が求まる
共通因子
（因子得点；因子量）
が求まる
誤差
（独自部分）
要因
因子負荷量
共通因子
$x_{A品}＝56$
ε品
ε雰
ε親
ε広
$x_{品}$
$x_{雰}$
$x_{親}$
$x_{広}$
$a_{品1}$=0.8
$a_{雰1}$
$a_{雰2}$
$a_{品2}$=0.2
$a_{親1}$
$a_{親2}$
$a_{広1}$
$a_{広2}$
f_1
f_2
$f_{A1}＝60$
$f_{A2}＝40$

データの数学的表現

1号店～n号店の各変量（要因）

店	品揃え（$x_{\text{品}}$）	雰囲気（$x_{\text{雰}}$）	親近感（$x_{\text{親}}$）	広々感（$x_{\text{広}}$）
1	$x_{1\text{品}}$	$x_{1\text{雰}}$	$x_{1\text{親}}$	$x_{1\text{広}}$
2	$x_{2\text{品}}$	$x_{2\text{雰}}$	$x_{2\text{親}}$	$x_{2\text{広}}$
⋮	⋮	⋮	⋮	⋮
k	$x_{k\text{品}}$	$x_{k\text{雰}}$	$x_{k\text{親}}$	$x_{k\text{広}}$
⋮	⋮	⋮	⋮	⋮
n	$x_{n\text{品}}$	$x_{n\text{雰}}$	$x_{n\text{親}}$	$x_{n\text{広}}$

k号店の場合

品揃えの評価 $\longrightarrow$ $x_{k\text{品}}$

雰囲気の評価 $\longrightarrow$ $x_{k\text{雰}}$

親近感の評価 $\longrightarrow$ $x_{k\text{親}}$

広々感の評価 $\longrightarrow$ $x_{k\text{広}}$

k号店

共通因子の因子量(因子得点)の数学的表現

1号店～n号店の各因子得点

店	共通因子1の因子得点(f_1)	共通因子2の因子得点(f_2)
1	f_{11}	f_{12}
2	f_{21}	f_{22}
⋮	⋮	⋮
k	f_{k1}	f_{k2}
⋮	⋮	⋮
n	f_{n1}	f_{n2}

k号店の因子得点

f_{k1} ･･･ k号店の持つ共通因子1の因子得点

f_{k2} ･･･ k号店の持つ共通因子2の因子得点

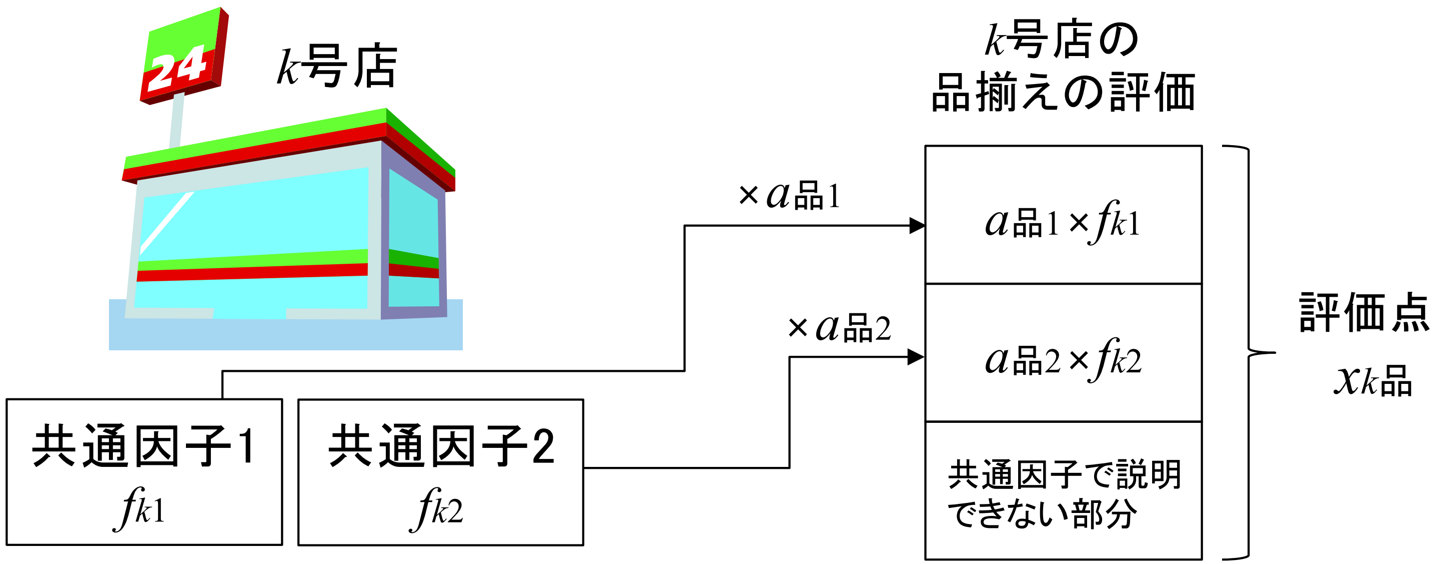
k号店

共通因子1	共通因子2
f_{k1}	f_{k2}

因子負荷量の数学的表現

因子負荷量：共通因子の各変量(要因)への影響度

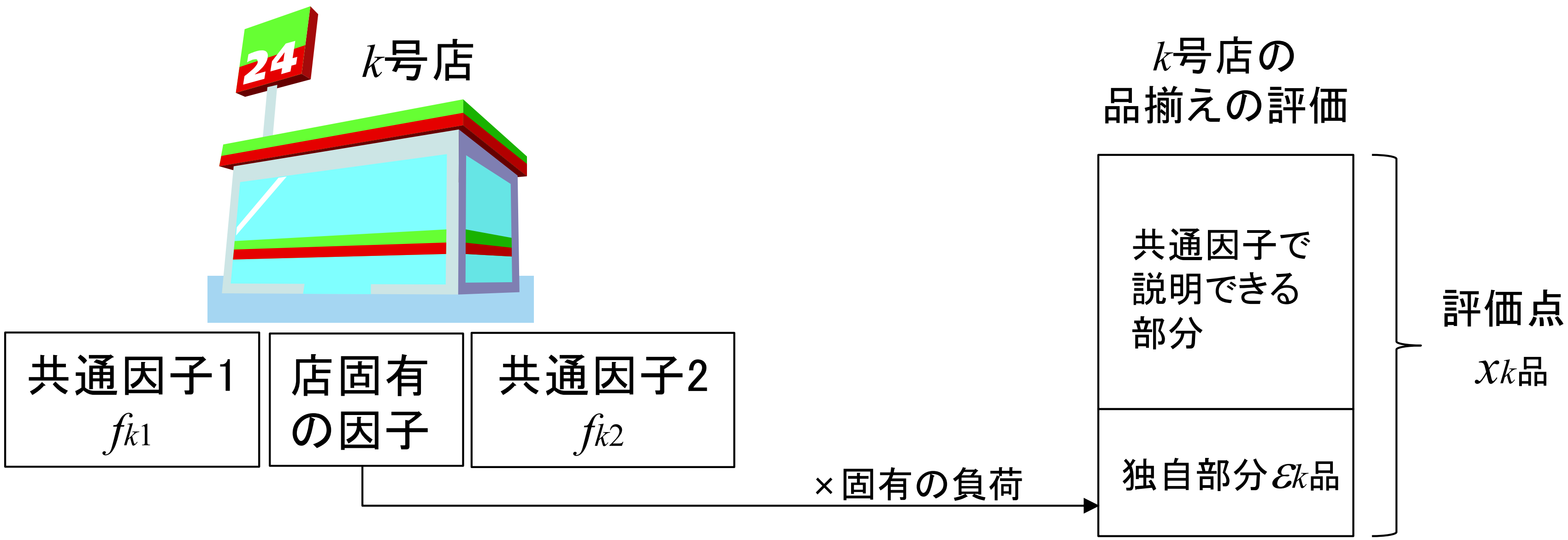
変量	共通因子1(f_1)に関わる 因子負荷量(a_1)	共通因子2(f_2)に関わる 因子負荷量(a_2)
品揃え($x_{品}$)	$a_{品1}$	$a_{品2}$
雰囲気($x_{雰}$)	$a_{雰1}$	$a_{雰2}$
親近感($x_{親}$)	$a_{親1}$	$a_{親2}$
広々感($x_{広}$)	$a_{広1}$	$a_{広2}$



誤差（独自部分）の数学的表現

1号店～n号店の共通因子では説明できない要素（独自部分）

店	品揃え（$x_{品}$）	雰囲気（$x_{雰}$）	親近感（$x_{親}$）	広々感（$x_{広}$）
1	$\varepsilon_{1品}$	$\varepsilon_{1雰}$	$\varepsilon_{1親}$	$\varepsilon_{1広}$
2	$\varepsilon_{2品}$	$\varepsilon_{2雰}$	$\varepsilon_{2親}$	$\varepsilon_{2広}$
⋮	⋮	⋮	⋮	⋮
k	$\varepsilon_{k品}$	$\varepsilon_{k雰}$	$\varepsilon_{k親}$	$\varepsilon_{k広}$
⋮	⋮	⋮	⋮	⋮
n	$\varepsilon_{n品}$	$\varepsilon_{n雰}$	$\varepsilon_{n親}$	$\varepsilon_{n広}$



因子分析の数学的表現

k号店の品揃えの評価点＝品揃えの共通因子1因子負荷量×k号店の共通因子1因子得点
＋品揃えの共通因子2因子負荷量×k号店の共通因子2因子得点
＋独自部分

$x_{k品} = a_{品1} \times f_{k1} + a_{品2} \times f_{k2} +$ 共通因子で説明できない部分（独自部分）
$= a_{品1} \times f_{k1} + a_{品2} \times f_{k2} + \varepsilon_{k品}$

k号店の因子分析の数学的表現

k号店の品揃えの評価点

$$x_{k品} = a_{品1} \times f_{k1} + a_{品2} \times f_{k2} + \varepsilon_{k品}$$

k号店の雰囲気の評価点

$$x_{k雰} = a_{雰1} \times f_{k1} + a_{雰2} \times f_{k2} + \varepsilon_{k雰}$$

k号店の親近感の評価点

$$x_{k親} = a_{親1} \times f_{k1} + a_{親2} \times f_{k2} + \varepsilon_{k親}$$

k号店の広々感の評価点

$$x_{k広} = a_{広1} \times f_{k1} + a_{広2} \times f_{k2} + \varepsilon_{k広}$$

因子分析の行列表現

k号店の因子分析の数学的表現

$$x_{k品} = a_{品1} \times f_{k1} + a_{品2} \times f_{k2} + \varepsilon_{k品}$$
$$x_{k雰} = a_{雰1} \times f_{k1} + a_{雰2} \times f_{k2} + \varepsilon_{k雰}$$
$$x_{k親} = a_{親1} \times f_{k1} + a_{親2} \times f_{k2} + \varepsilon_{k親}$$
$$x_{k広} = a_{広1} \times f_{k1} + a_{広2} \times f_{k2} + \varepsilon_{k広}$$

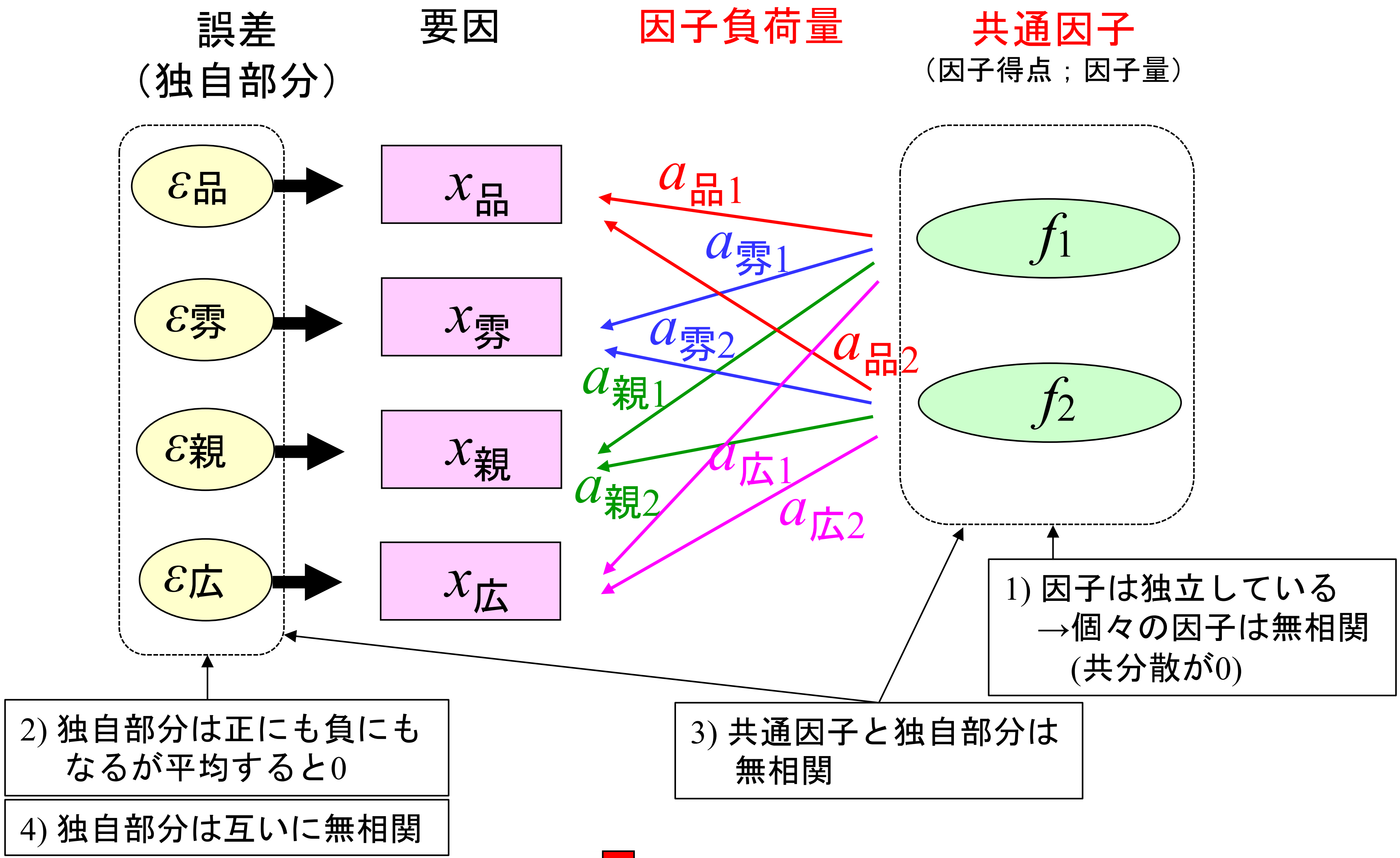
$1 \sim n$号店の因子分析の数学的表現

$$\boldsymbol{X} = \begin{bmatrix} x_{1品} & x_{1雰} & x_{1親} & x_{1広} \\ x_{2品} & x_{2雰} & x_{2親} & x_{2広} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n品} & x_{n雰} & x_{n親} & x_{n広} \end{bmatrix} \qquad \boldsymbol{F} = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \\ \vdots & \vdots \\ f_{n1} & f_{n2} \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} a_{品1} & a_{品2} \\ a_{雰1} & a_{雰2} \\ a_{親1} & a_{親2} \\ a_{広1} & a_{広2} \end{bmatrix} \qquad \boldsymbol{E} = \begin{bmatrix} \varepsilon_{1品} & \varepsilon_{1雰} & \varepsilon_{1親} & \varepsilon_{1広} \\ \varepsilon_{2品} & \varepsilon_{2雰} & \varepsilon_{2親} & \varepsilon_{2広} \\ \vdots & \vdots & \vdots & \vdots \\ \varepsilon_{n品} & \varepsilon_{n雰} & \varepsilon_{n親} & \varepsilon_{n広} \end{bmatrix}$$

因子モデル $\boldsymbol{X} = \boldsymbol{F}\boldsymbol{A}^t + \boldsymbol{E}$

因子をどの様に算出するのか?



反復推定法による収束計算を行ない因子負荷量、因子得点を算出する

データの具体例（データの入力）

1号店～10号店の各変量（要因）

店＼要因	品揃え（$x_{品}$）	雰囲気（$x_{雰}$）	親近感（$x_{親}$）	広々感（$x_{広}$）
1	20	20	30	70
2	15	25	25	15
3	60	70	60	30
4	70	50	60	40
5	30	35	55	80
6	50	85	85	80
7	80	65	70	30
8	30	65	75	80
9	60	60	50	40
10	25	25	50	25

因子分析では最初にデータの標準化が行われる

1号店～10号店の各変量（要因，標準化後）　$\boldsymbol{X} = \boldsymbol{F}\boldsymbol{A}^t + \boldsymbol{E}$

店＼要因	品揃え（$x_{品}$）	雰囲気（$x_{雰}$）	親近感（$x_{親}$）	広々感（$x_{広}$）
1	-1.051	-1.334	-1.392	0.817
2	-1.270	-1.112	-1.660	-1.323
3	0.701	0.889	0.214	-0.740
4	1.139	0.000	0.214	-0.350
5	-0.613	-0.667	-0.054	1.207
6	0.263	1.557	1.553	1.207
7	1.577	0.667	0.750	-0.740
8	-0.613	0.667	1.017	1.207
9	0.701	0.445	-0.321	-0.350
10	-0.832	-1.112	-0.321	-0.934

因子分析では最初にデータの標準化が行われる

因子負荷量の具体例（出力結果）

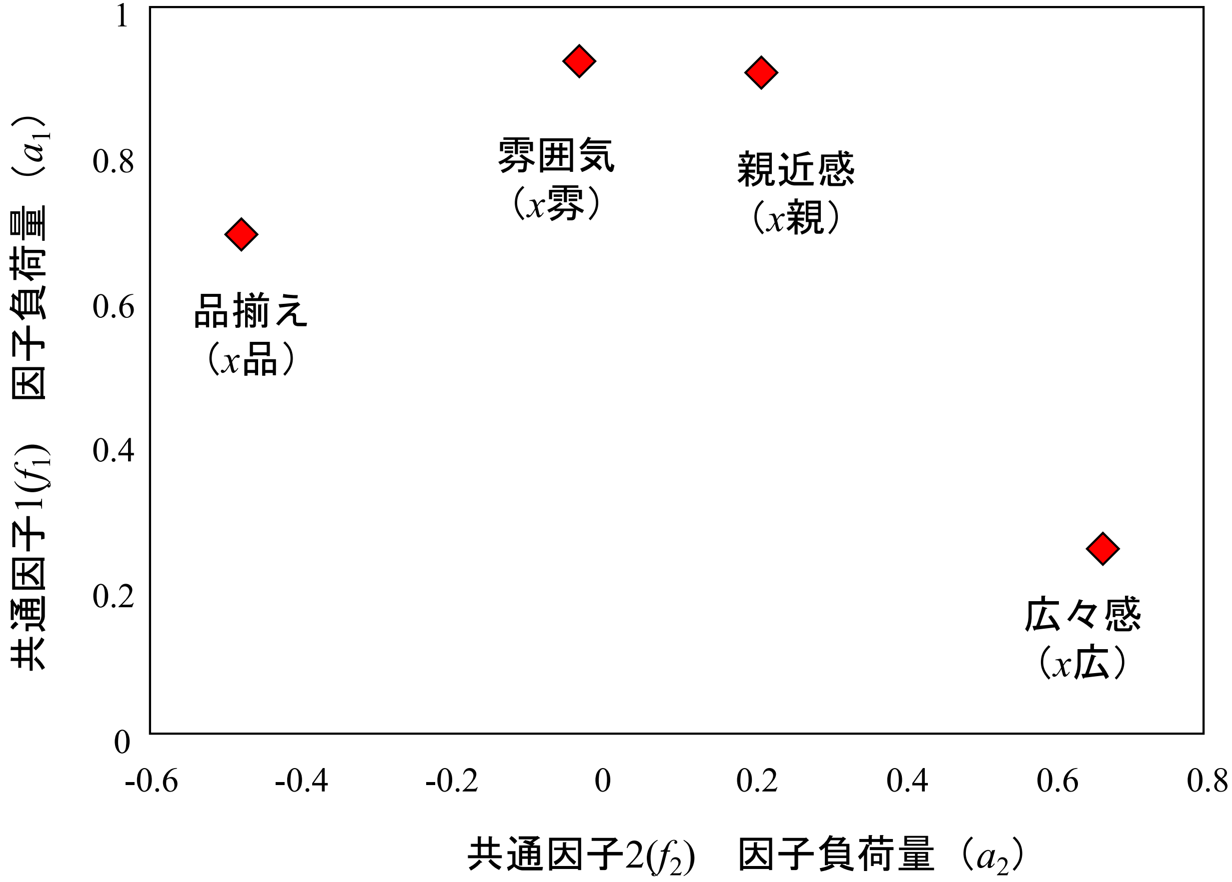
因子負荷量　$\boldsymbol{X} = \boldsymbol{F}\boldsymbol{A}^{t} + \boldsymbol{E}$

変量	共通因子1(f_1)に関わる 因子負荷量(a_1)	共通因子2(f_2)に関わる 因子負荷量(a_2)
品揃え($x_{品}$)	0.68	-0.48
雰囲気($x_{雰}$)	0.92	-0.03
親近感($x_{親}$)	0.91	0.21
広々感($x_{広}$)	0.25	0.67

因子負荷量：共通因子の各変量(要因)への影響度

出力結果の解釈

変量	共通因子1(f_1)に関わる 因子負荷量（a_1）	共通因子2(f_2)に関わる 因子負荷量（a_2）
品揃え（$x_{品}$）	0.68	-0.48
雰囲気（$x_{雰}$）	0.92	-0.03
親近感（$x_{親}$）	0.91	0.21
広々感（$x_{広}$）	0.25	0.67



共通因子の解釈

共通因子1 (f_1)→
店長の工夫という要因
（ソフト）

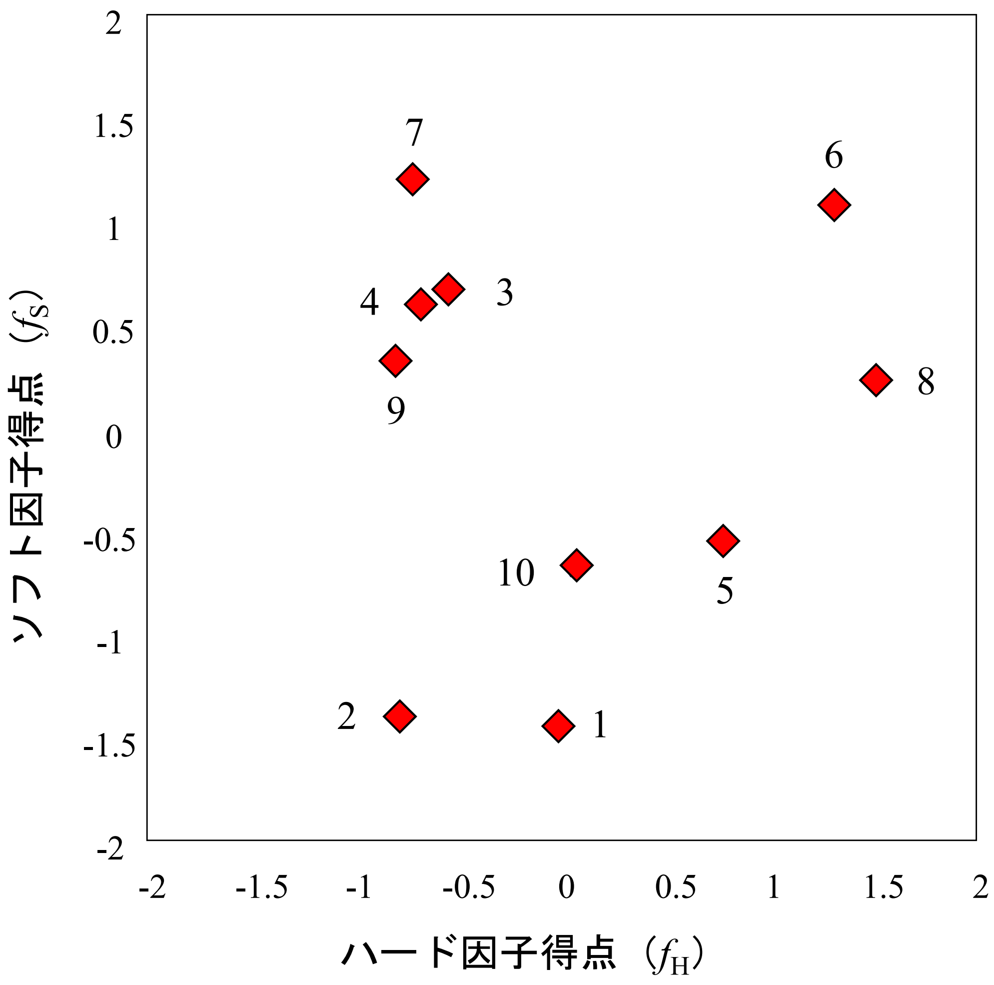
共通因子2 (f_2)→
地理的・物理的条件という要因
（ハード）

共通因子の因子得点の具体例

1号店～10号店の各因子得点　$\boldsymbol{X} = \boldsymbol{F}\boldsymbol{A}^t + \boldsymbol{E}$

店	ソフト 因子得点（f_S）	ハード 因子得点（f_H）
1	-1.4513	-0.0326
2	-1.4035	-0.7977
3	0.6841	-0.5743
4	0.6027	-0.6981
5	-0.5664	0.7809
6	1.0595	1.3198
7	1.2218	-0.7381
8	0.2334	1.5272
9	0.3074	-0.8290
10	-0.6896	0.0420

店	ソフト因子得点 (f_S)	ハード因子得点 (f_H)
1	-1.4513	-0.0326
2	-1.4035	-0.7977
3	0.6841	-0.5743
4	0.6027	-0.6981
5	-0.5664	0.7809
6	1.0595	1.3198
7	1.2218	-0.7381
8	0.2334	1.5272
9	0.3074	-0.8290
10	-0.6896	0.0420



因子得点から因子の保有量がわかる

ex.

1号店はハード因子得点は平均的

しかしソフト因子得点は足りない

⬇

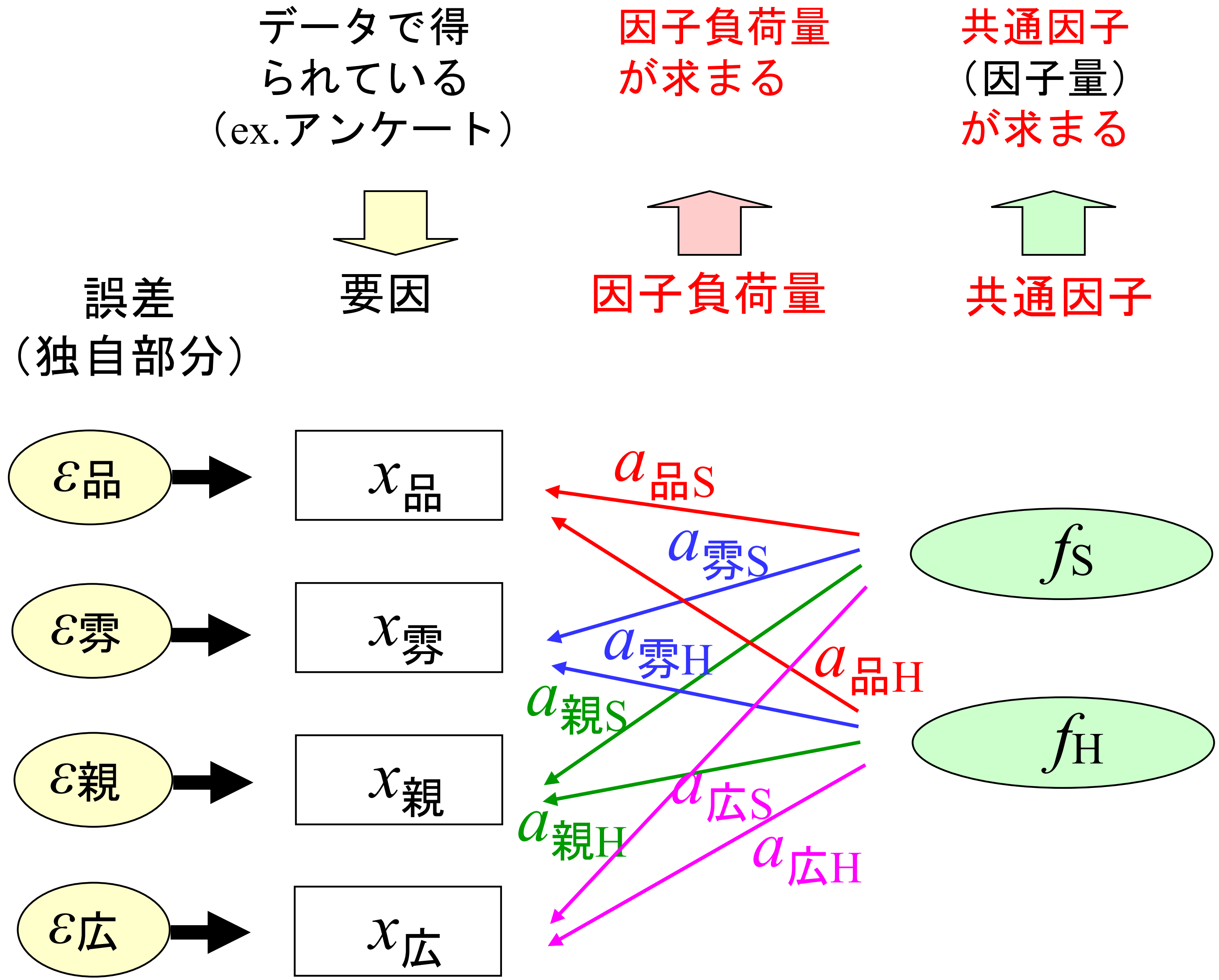
経営努力を促す

1号店～10号店の要因（独自成分なし）　$\boldsymbol{X} = \boldsymbol{FA^t} + \boldsymbol{E}$

要因 / 店	品揃え（$x_{品}$）	雰囲気（$x_{雰}$）	親近感（$x_{親}$）	広々感（$x_{広}$）
1	-0.971	-1.334	-1.328	-0.385
2	-0.571	-1.267	-1.445	-0.885
3	0.741	0.647	0.502	-0.214
4	0.745	0.575	0.402	-0.317
5	-0.760	-0.545	-0.351	0.382
6	0.087	0.935	1.241	1.149
7	1.185	1.146	0.957	-0.189
8	-0.574	0.169	0.533	1.082
9	0.607	0.308	0.106	-0.479
10	-0.489	-0.636	-0.619	-0.144

1号店～10号店の誤差（独自成分）　　$\boldsymbol{X} = \boldsymbol{F}\boldsymbol{A}^{t} + \boldsymbol{E}$

店＼要因	品揃え（$x_{品}$）	雰囲気（$x_{雰}$）	親近感（$x_{親}$）	広々感（$x_{広}$）
1	-0.080	0.000	-0.064	1.202
2	-0.699	0.155	-0.215	-0.438
3	-0.040	0.243	-0.288	-0.526
4	0.394	-0.575	-0.188	-0.033
5	0.147	-0.123	0.298	0.825
6	0.176	0.621	0.311	0.058
7	0.392	-0.479	-0.207	-0.550
8	-0.039	0.498	0.484	0.125
9	0.094	0.137	-0.427	0.128
10	-0.343	-0.476	0.297	-0.790



データで得
られている
（ex.アンケート）
因子負荷量
が求まる
共通因子
（因子量）
が求まる
誤差
（独自部分）
要因
因子負荷量
共通因子
ε品
ε雰
ε親
ε広
x品
x雰
x親
x広
a品S
a雰S
a雰H
a品H
a親S
a親H
a広S
a広H
fS
fH

因子の解釈の例

高齢者の転倒事故が多く見られる住宅空間についての調査結果

要因

No.	寝室	居間	階段	ベランダ	浴室	トイレ	食堂	玄関	庭	廊下
1	2	1	3	3	4	4	3	4	3	2
2	1	1	3	4	5	4	5	5	3	3
3	2	2	3	3	2	4	4	5	1	4
4	3	1	3	3	4	4	4	3	3	4
5	3	1	3	3	3	3	3	3	3	3
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
78	1	3	4	3	4	4	5	1	3	4
79	1	1	2	3	3	3	3	4	3	4
80	3	1	3	1	1	1	3	2	1	2

要因

No.	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{110}
2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{210}
3	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}	x_{37}	x_{38}	x_{39}	x_{310}
4	x_{41}	x_{42}	x_{42}	x_{44}	x_{45}	x_{46}	x_{47}	x_{48}	x_{49}	x_{410}
5	x_{51}	x_{52}	x_{53}	x_{54}	x_{55}	x_{56}	x_{57}	x_{58}	x_{59}	x_{510}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
78										
79										
80										

SPSSでの因子分析結果

要因　　因子負荷量　　共通因子

	因子	
	1	2
浴室	.858	-.098
食堂	.844	.039
トイレ	.744	.023
廊下	.634	.168
庭	.602	.029
玄関	.063	.818
ベランダ	.292	.459
階段	-.101	.372
居間	.161	-.091
寝室	-.311	.211

要因　　因子負荷量　　共通因子

	因子	
	f_1	f_2
x_1	a_{11}	a_{12}
x_2	a_{21}	a_{22}
x_3	a_{31}	a_{32}
x_4	a_{41}	a_{42}
x_5	a_{51}	a_{52}
x_6	a_{61}	a_{62}
x_7	a_{71}	a_{72}
x_8	a_{81}	a_{82}
x_9	a_{91}	a_{92}
x_{10}	a_{101}	a_{102}

SPSSでの因子分析結果

要因　因子負荷量　共通因子

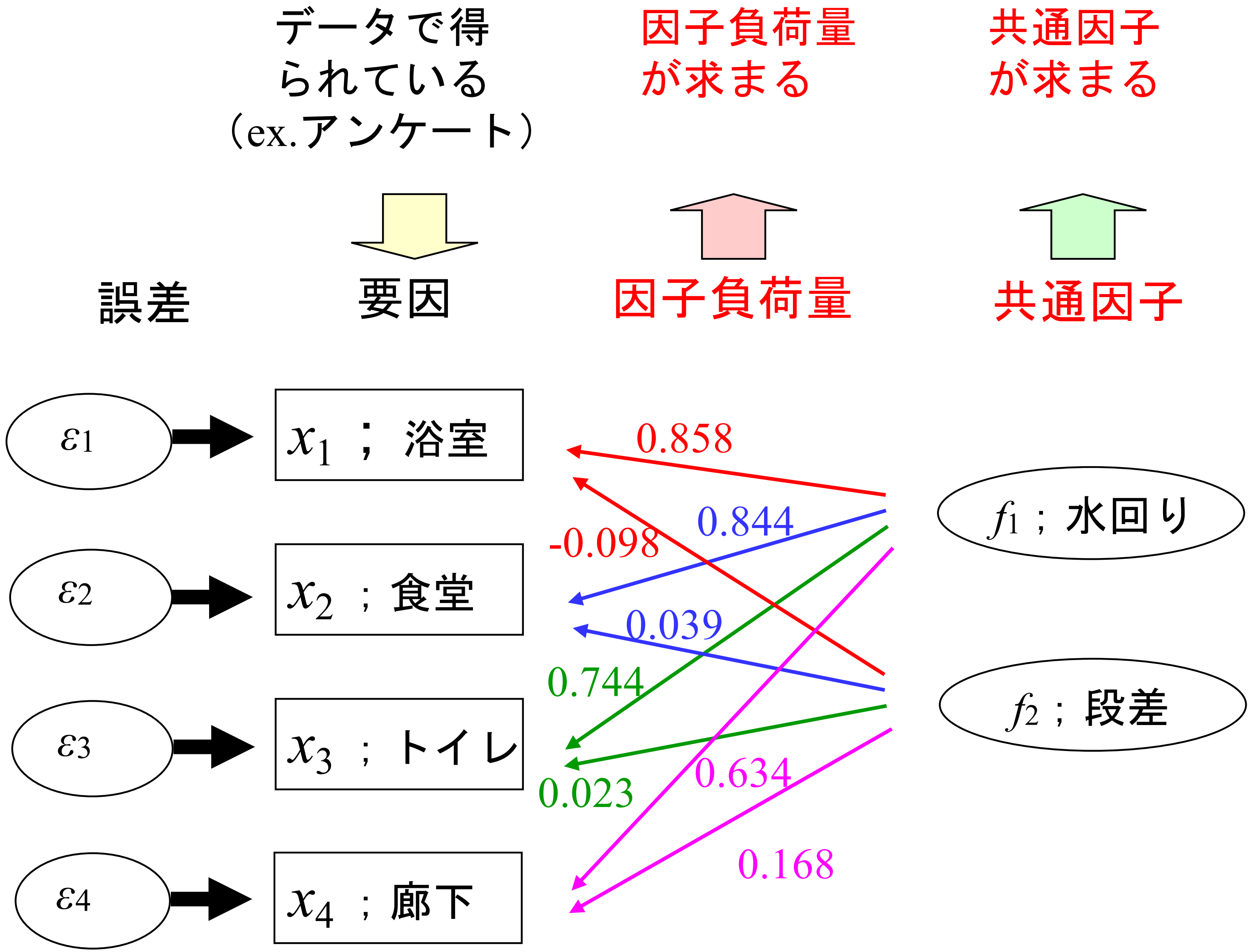
	因子	
	1	2
浴室	.858	-.098
食堂	.844	.039
トイレ	.744	.023
廊下	.634	.168
庭	.602	.029
玄関	.063	.818
ベランダ	.292	.459
階段	-.101	.372
居間	.161	-.091
寝室	-.311	.211

共通因子の解釈

第１因子は
浴室、食堂、トイレ
といった変数の因子負荷が大きいので
“水まわり”

第２因子は
玄関、ベランダ、階段
といった変数の因子負荷が大きいので
“段差のあるところ”

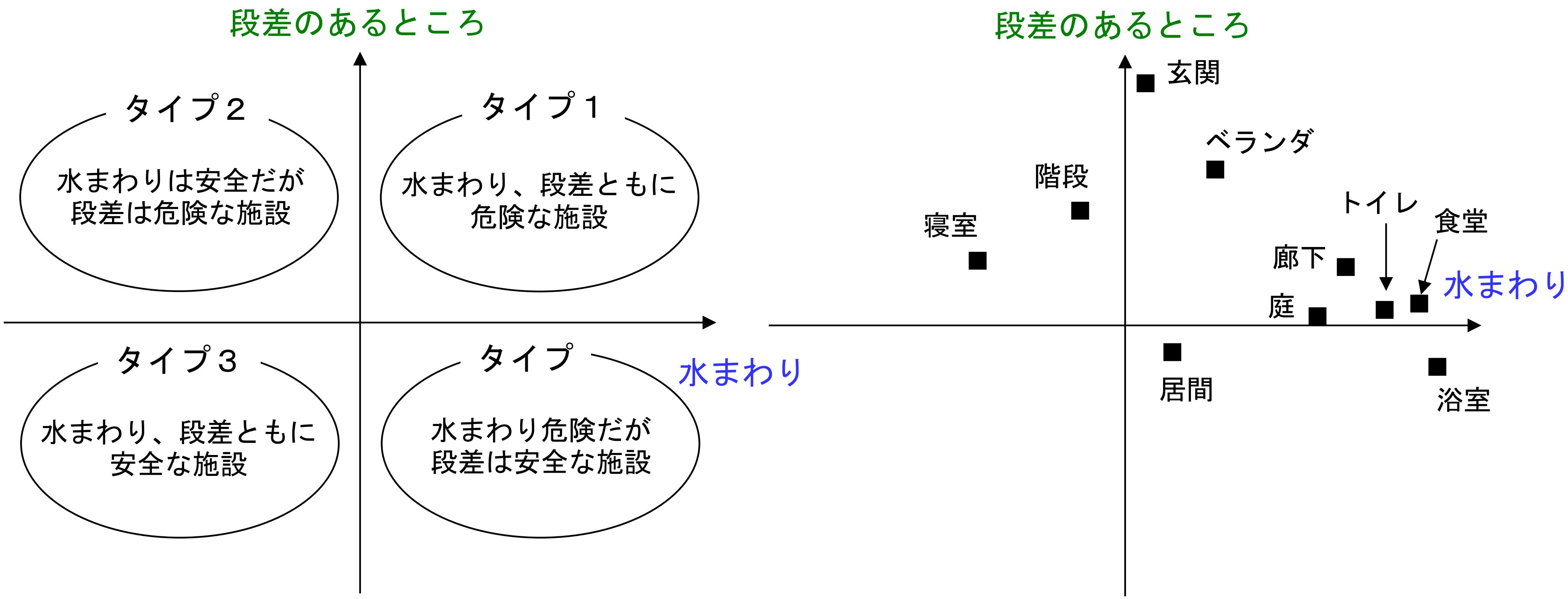
と名付けられそう。



データで得
られている
（ex.アンケート）
因子負荷量
が求まる
共通因子
が求まる
誤差
要因
因子負荷量
共通因子
ε1
ε2
ε3
ε4
x1 ； 浴室
x2 ； 食堂
x3 ； トイレ
x4 ； 廊下
f1 ； 水回り
f2 ； 段差
0.858
-0.098
0.844
0.039
0.744
0.023
0.634
0.168

どのような住宅空間が高齢者にとって危険なのかを
因子分析でタイプ分けをすると

共通因子として
第１因子：水まわり
第２因子：段差のあるところ
が抽出される



第１因子と第２因子の散布図

因子分析

①因子分析とは何か。
要因が３個（x_1, x_2, x_3）、共通因子が２個（f_1, f_2）の場合を例にとってパス図で示せ。

②大学生が入社以前に身につけておくべきだと思われる知識・技能に関するアンケート結果の因子分析を行った結果、右に示す表の結果を得た。
第１因子と第２因子に名前を付け散布図で示せ。

	因子	
	1	2
一般常識	0.778	−0.162
礼儀作法	0.767	−0.055
文章力	0.474	0.716
企画力	−0.094	0.739
法律知識	−0.173	0.546
パソコン操作	0.158	0.099
実務経験	0.100	0.116
語学力	0.074	0.212
各種資格	0.026	−0.083