

Rによる因子分析

奥村太一 (2016.2.1)

概要

右のデータは、200人の生徒について得られた“国語”、“英語”、“社会”、“数学”、“理科”、“技術”それぞれの成績を5段階評価で示したものである。¹ここでは、Rを用いた因子分析により、これら6教科の成績を左右する潜在的な特性について探索することにする。

	A	B	C	D	E	F	G
1	生徒	国語	英語	社会	数学	理科	技術
2	1	4	4	4	4	3	5
3	2	4	4	4	4	5	5
4	3	3	3	4	4	4	4
5	4	3	3	3	3	3	2
6	5	3	3	2	2	3	2

データの読み込み

まずは、下のプログラムによってデータを読み込む。関数 `read.csv()` 内の引数 `row.names=1` は、このデータの1列目は変数ではなく生徒を識別する番号であることを宣言している。また `head()` は、読み込んだデータの先頭数行だけを表示させる関数である。

```
data01 <- read.csv(file="/Users/Name/Desktop/factor.csv",header=T,
                  fileEncoding="Shift-JIS", row.names=1)
head(data01)
```

分析の実行

相関行列の算出と因子数の決定

Rには、相関行列を算出する関数 `cor()` がある。これを用いて、`data01`に含まれる6つの変数間の相関行列を算出し、それを `cor01` に代入し表示させてみよう。

`cor()` データの相関行列を算出する関数

```
cor01 <- cor(data01); cor01
```

すると、次のように相関行列が返されるはずである。

	国語	英語	社会	数学	理科	技術
国語	1.0000000	0.4158982	0.4836725	0.2505441	0.2907936	0.2225323
英語	0.4158982	1.0000000	0.4980821	0.2649821	0.2946882	0.3051505
社会	0.4836725	0.4980821	1.0000000	0.2945884	0.4020571	0.3465945
数学	0.2505441	0.2649821	0.2945884	1.0000000	0.5766291	0.5583059
理科	0.2907936	0.2946882	0.4020571	0.5766291	1.0000000	0.5253087
技術	0.2225323	0.3051505	0.3465945	0.5583059	0.5253087	1.0000000

次に、共通因子の数を定める手がかりとして、この相関行列の固有値を算出し、スクリーンプロットを描いてみる。Rには固有値と固有ベクトルを算出する関数 `eigen()` があり、これによって返された結果の `values` に固有値が格納されている。

`eigen()` 行列の固有値および固有ベクトルを算出する関数。固有値は `values` に格納される。

¹ 最初の5名分だけを表示してある。なお、このデータは <http://www.juen.ac.jp/lab/okumura/data.html> から“factor.csv”としてダウンロード可能である (リンクを右クリックして保存)。

これを用いて、相関行列 cor01 の固有値を eigen01 に代入し、表示させてみよう。

```
eigen01 <- eigen(cor01)$values; eigen01
```

すると、以下のように6つの固有値が得られるはずである。第1固有値、第2固有値の2つが1を越えていることがわかる。

```
[1] 2.9199728 1.1337679 0.5975924 0.5004401 0.4595162 0.3887106
```

この固有値の減衰状況をスクリープロットによって調べてみよう。Rには plot() という描画関数があり、これによってスクリープロットは簡単に描くことができる。plot() には多くの引数を用意されているが、ここで用いているのは以下の通りである。

plot()	与えられたデータをもとに描画を行う関数	
	type	プロットの種類を指定する。“p”なら点のみ、“l”なら線のみ、“b”は点と線と両方、など。
	main	プロットのタイトルを“”でくくって指定する。
	xlab	x軸の名前を“”でくくって指定する。
	ylab	y軸の名前を“”でくくって指定する。

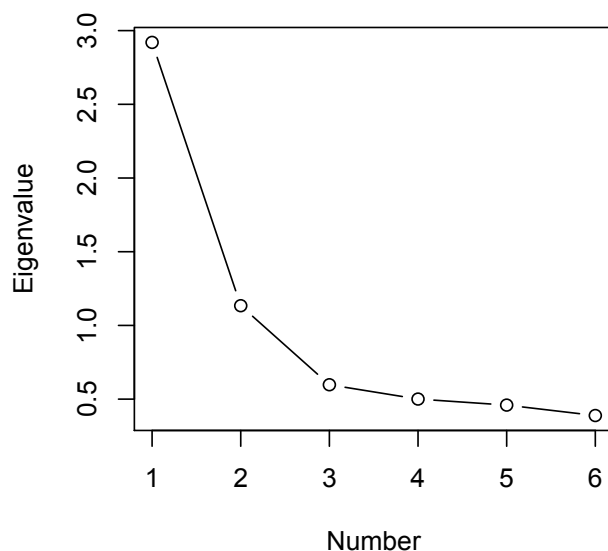
eigen01 に格納された固有値をプロットするには、以下のコマンドを実行すればよい。

```
plot(eigen01, type="b", main="Scree Plot", xlab="Number", ylab="Eigenvalue")
```

これによって得られたスクリープロットが下の図である。第3固有値以降は、減衰が穏やかになっている様子が見て取れる。

ここでは、1より大きい固有値の数とスクリープロットにおける固有値の減衰状況から、共通因子の数を2として因子分析を行うこととする。

Scree Plot



因子分析の実行

Rには、因子分析を行うための関数 factanal() が標準で装備されている。この関数では、分析対象となるデータと、共通因子の数、因子の回転法を指定するだけで因子分析を実行できる。

ただし、初期解の算出は最尤法のみとなっており、因子の回転法としてはバリマックス法（直交回転）とプロマックス法（斜交回転）しか選択できない（回転する前の初期解を得ることもできる）。

factanal()	因子分析を行う関数	
	x	データ行列を指定する。
	factors	共通因子の数を指定する。
	rotation	因子の回転法を“varimax”もしくは“promax”から選択する。回転させない場合は“none”と指定。

今、回転なし（初期解）、バリマックス回転とプロマックス回転による結果をそれぞれ fit00、fit01、fit02として代入しておく。この結果は、関数 print() により得ることができる。なお、引数 cutoff=0 は（値の小さいものも含め）全ての因子負荷の値を表示させる指示である。

```
fit00 <- factanal(x=data01, factors=2, rotation="none")
print(fit00, cutoff=0)
fit01 <- factanal(x=data01, factors=2, rotation="varimax")
print(fit01, cutoff=0)
fit02 <- factanal(x=data01, factors=2, rotation="promax")
print(fit02, cutoff=0)
```

上記のプログラムを実行すると、初期解、バリマックス解、プロマックス解が以下のように得られる。ただし、factanal() では共通性の値は出力されない。そのかわり、独自性 (=1-共通性) の値が Uniquenesses として表示される。因子の回転によって因子負荷と寄与の値は変化するが、独自性の値は変化しない（ということは共通性の値も変化しない）ことも確認しておこう。

回転後の因子負荷行列を見る限り、第1因子（Factor1）を“理系学力”、第2因子（Factor2）を“文系学力”と命名できそうである。バリマックス解よりもプロマックス解の方が単純構造が満たされており、因子間相関も0.57と正の強い値が得られていることから、プロマックス解を採用するのが妥当であろう。

初期解	バリマックス解	プロマックス解																																																																																																																								
Loadings: <table border="1"> <thead> <tr> <th></th> <th>Factor1</th> <th>Factor2</th> </tr> </thead> <tbody> <tr><td>国語</td><td>0.510</td><td>0.362</td></tr> <tr><td>英語</td><td>0.545</td><td>0.344</td></tr> <tr><td>社会</td><td>0.656</td><td>0.420</td></tr> <tr><td>数学</td><td>0.704</td><td>-0.363</td></tr> <tr><td>理科</td><td>0.715</td><td>-0.199</td></tr> <tr><td>技術</td><td>0.670</td><td>-0.238</td></tr> </tbody> </table> <table border="1"> <thead> <tr> <th></th> <th>Factor1</th> <th>Factor2</th> </tr> </thead> <tbody> <tr><td>SS loadings</td><td>2.445</td><td>0.653</td></tr> <tr><td>Proportion Var</td><td>0.408</td><td>0.109</td></tr> <tr><td>Cumulative Var</td><td>0.408</td><td>0.516</td></tr> </tbody> </table> <table border="1"> <thead> <tr> <th></th> <th>Factor1</th> <th>Factor2</th> </tr> </thead> <tbody> <tr><td>寄与</td><td></td><td></td></tr> <tr><td>寄与率</td><td></td><td></td></tr> <tr><td>累積寄与率</td><td></td><td></td></tr> </tbody> </table>		Factor1	Factor2	国語	0.510	0.362	英語	0.545	0.344	社会	0.656	0.420	数学	0.704	-0.363	理科	0.715	-0.199	技術	0.670	-0.238		Factor1	Factor2	SS loadings	2.445	0.653	Proportion Var	0.408	0.109	Cumulative Var	0.408	0.516		Factor1	Factor2	寄与			寄与率			累積寄与率			Loadings: <table border="1"> <thead> <tr> <th></th> <th>Factor1</th> <th>Factor2</th> </tr> </thead> <tbody> <tr><td>国語</td><td>0.163</td><td>0.604</td></tr> <tr><td>英語</td><td>0.202</td><td>0.612</td></tr> <tr><td>社会</td><td>0.238</td><td>0.742</td></tr> <tr><td>数学</td><td>0.774</td><td>0.169</td></tr> <tr><td>理科</td><td>0.678</td><td>0.302</td></tr> <tr><td>技術</td><td>0.668</td><td>0.244</td></tr> </tbody> </table> <table border="1"> <thead> <tr> <th></th> <th>Factor1</th> <th>Factor2</th> </tr> </thead> <tbody> <tr><td>SS loadings</td><td>1.629</td><td>1.469</td></tr> <tr><td>Proportion Var</td><td>0.272</td><td>0.245</td></tr> <tr><td>Cumulative Var</td><td>0.272</td><td>0.516</td></tr> </tbody> </table>		Factor1	Factor2	国語	0.163	0.604	英語	0.202	0.612	社会	0.238	0.742	数学	0.774	0.169	理科	0.678	0.302	技術	0.668	0.244		Factor1	Factor2	SS loadings	1.629	1.469	Proportion Var	0.272	0.245	Cumulative Var	0.272	0.516	Loadings: <table border="1"> <thead> <tr> <th></th> <th>Factor1</th> <th>Factor2</th> </tr> </thead> <tbody> <tr><td>国語</td><td>-0.035</td><td>0.645</td></tr> <tr><td>英語</td><td>0.007</td><td>0.641</td></tr> <tr><td>社会</td><td>0.001</td><td>0.778</td></tr> <tr><td>数学</td><td>0.836</td><td>-0.081</td></tr> <tr><td>理科</td><td>0.675</td><td>0.109</td></tr> <tr><td>技術</td><td>0.685</td><td>0.044</td></tr> </tbody> </table> <table border="1"> <thead> <tr> <th></th> <th>Factor1</th> <th>Factor2</th> </tr> </thead> <tbody> <tr><td>SS loadings</td><td>1.624</td><td>1.452</td></tr> <tr><td>Proportion Var</td><td>0.271</td><td>0.242</td></tr> <tr><td>Cumulative Var</td><td>0.271</td><td>0.513</td></tr> </tbody> </table> Factor Correlations: <table border="1"> <thead> <tr> <th></th> <th>Factor1</th> <th>Factor2</th> </tr> </thead> <tbody> <tr><td>Factor1</td><td>1.000</td><td>0.571</td></tr> <tr><td>Factor2</td><td>0.571</td><td>1.000</td></tr> </tbody> </table>		Factor1	Factor2	国語	-0.035	0.645	英語	0.007	0.641	社会	0.001	0.778	数学	0.836	-0.081	理科	0.675	0.109	技術	0.685	0.044		Factor1	Factor2	SS loadings	1.624	1.452	Proportion Var	0.271	0.242	Cumulative Var	0.271	0.513		Factor1	Factor2	Factor1	1.000	0.571	Factor2	0.571	1.000
	Factor1	Factor2																																																																																																																								
国語	0.510	0.362																																																																																																																								
英語	0.545	0.344																																																																																																																								
社会	0.656	0.420																																																																																																																								
数学	0.704	-0.363																																																																																																																								
理科	0.715	-0.199																																																																																																																								
技術	0.670	-0.238																																																																																																																								
	Factor1	Factor2																																																																																																																								
SS loadings	2.445	0.653																																																																																																																								
Proportion Var	0.408	0.109																																																																																																																								
Cumulative Var	0.408	0.516																																																																																																																								
	Factor1	Factor2																																																																																																																								
寄与																																																																																																																										
寄与率																																																																																																																										
累積寄与率																																																																																																																										
	Factor1	Factor2																																																																																																																								
国語	0.163	0.604																																																																																																																								
英語	0.202	0.612																																																																																																																								
社会	0.238	0.742																																																																																																																								
数学	0.774	0.169																																																																																																																								
理科	0.678	0.302																																																																																																																								
技術	0.668	0.244																																																																																																																								
	Factor1	Factor2																																																																																																																								
SS loadings	1.629	1.469																																																																																																																								
Proportion Var	0.272	0.245																																																																																																																								
Cumulative Var	0.272	0.516																																																																																																																								
	Factor1	Factor2																																																																																																																								
国語	-0.035	0.645																																																																																																																								
英語	0.007	0.641																																																																																																																								
社会	0.001	0.778																																																																																																																								
数学	0.836	-0.081																																																																																																																								
理科	0.675	0.109																																																																																																																								
技術	0.685	0.044																																																																																																																								
	Factor1	Factor2																																																																																																																								
SS loadings	1.624	1.452																																																																																																																								
Proportion Var	0.271	0.242																																																																																																																								
Cumulative Var	0.271	0.513																																																																																																																								
	Factor1	Factor2																																																																																																																								
Factor1	1.000	0.571																																																																																																																								
Factor2	0.571	1.000																																																																																																																								
Uniquenesses: 国語 英語 社会 数学 理科 技術 0.609 0.584 0.393 0.372 0.449 0.494	独自性 (= 1-共通性)																																																																																																																									

因子分析の実行 (psychパッケージを利用する場合)

Rにデフォルトで用意されている factanal() では、初期解の算出法や因子の回転法に関して選択の幅がかなり限られている。一方、心理データの統計分析に特化したパッケージ psych に装備されている関数 fa() を用いることで、より様々なオプションを選択することが可能となる。ただし、psych パッケージはデフォルトではRに入っていないので、install.packages("psych") によってダウンロード・インストールし、使用時に library("psych") により読み込む必要がある。

fa()	psychパッケージにおいて因子分析を行う関数	
r	データ行列を指定する。	
nfactors	共通因子の数を指定する。	
rotate	因子の回転法を指定する。factanal()と同じ“none”、“varimax”、“promax”以外に、“quartimax”(クオーティマックス法)、“equamax”(エカマックス法)などの直交回転、“oblimin”(オブリミン法)などの斜交回転を選択できる。	
fm	初期解の算出法を指定する。“ml”(最尤法)以外に、“minres”(最小2乗法)などが選択できる。	

factanal() を用いた場合と同じように、因子数を2とし、最尤法によって得られた初期解にバリマックス法、プロマックス法を適用するとする。これらの分析結果をそれぞれ fit11 と fit12 に格納し表示させるには、以下のようにすればよい。

```
library("psych")
fit11 <- fa(r=data01, nfactors=2, rotate="varimax", fm="ml")
print(fit11)
fit12 <- fa(r=data01, nfactors=2, rotate="promax", fm="ml")
print(fit12)
```

結果を一部抜粋して以下にまとめておく。最尤法により初期解を推定した場合は、fa() のアウトプットでは因子名は ML1, ML2, … となり、その下に因子負荷が並んでいる。因子負荷の右側の3列は、それぞれ共通性 (h2)、独自性 (u2)、複雑性 (com)² である。

寄与については、寄与率と累積寄与率に加え、説明率（その因子の寄与/寄与の和）と累積説明率（説明率を第1因子から足し合わせたもの）が報告されている。

データによっては、初期解の算出法や回転法を変えることで結果が大きく変わる場合がある。ここで用いたデータではどうだろうか。各自で試してみてください。

バリマックス解					プロマックス解						
	ML1	ML2	h2	u2	com		ML1	ML2	h2	u2	com
国語	0.16	0.60	0.39	0.61	1.1	国語	-0.03	0.64	0.39	0.61	1.0
英語	0.20	0.61	0.42	0.58	1.2	英語	0.01	0.64	0.42	0.58	1.0
社会	0.24	0.74	0.61	0.39	1.2	社会	0.00	0.78	0.61	0.39	1.0
数学	0.77	0.17	0.63	0.37	1.1	数学	0.84	-0.08	0.63	0.37	1.0
理科	0.68	0.30	0.55	0.45	1.4	理科	0.67	0.11	0.55	0.45	1.1
技術	0.67	0.24	0.51	0.49	1.3	技術	0.69	0.04	0.51	0.49	1.0
			ML1	ML2					ML1	ML2	
SS loadings			1.63	1.47		SS loadings			1.63	1.46	
Proportion Var			0.27	0.24		Proportion Var			0.27	0.24	
Cumulative Var			0.27	0.52		Cumulative Var			0.27	0.52	
Proportion Explained			0.53	0.47		Proportion Explained			0.53	0.47	
Cumulative Proportion			0.53	1.00		Cumulative Proportion			0.53	1.00	
			With factor correlations of						ML1 ML2		
			ML1	ML2				ML1	ML2		
			ML1	1.00	0.57			ML1	1.00	0.57	
			ML2	0.57	1.00			ML2	0.57	1.00	

因子負荷 共通性 独自性 複雑性

寄与
寄与率
累積寄与率
説明率
累積説明率

因子間相関

² 複雑性 (complexity) は、単純構造の程度を観測変数ごとに評価する指標である。単一の因子のみから負荷を得ていれば最小値1を取り、全ての因子から同じ負荷を得ていれば最大値（因子数に等しくなる）を取る。つまり、複雑性の値が1に近い項目ほど、単純構造の成立に貢献していると判断できる。