

12日目：因子分析（2）

さて、昨日はいろいろと分析を試していただけたでしょうか？ 因子分析は、かなりの試行錯誤を必要とするので、じっくりと構えて取り組む必要があります。

投入する変数を削除することにも挑戦した人はわかるでしょうが、以下のような変数をまとめておく作業は、ここで結構役立ちます。ここをちょっと変えるだけで、投入する変数を簡単に(?) 変えられるからです。

```
xtb <- c("b1", "b2", "b3", "b4", "b5", "b6", "b7", "b8", "b9", "b10", "b11",  
"b12", "b13", "b14", "b15", "b16", "b17", "b18", "b19", "b20" )  
xb <- x[xtb]
```

今日は、因子分析結果として表示されるものの説明と、結果をエクセルに移しての加工について説明します。

さて、因子分析の結果ですが、SPSS よりも多くの情報を示してくれます。私もよくわかっていない部分も少なくないですが、とりあえず、「お約束」のあたりを紹介しておきます。

```
f1 <- fa(xb, nfactors=3, fm="ml", rotate="promax")  
print(f1, sort=TRUE, digit=3)
```

とりあえず、以上のように、抽出因子数は3、抽出方法は最尤法、プロマックス回転をかけ、パターンの大きさによるソートをかけた結果の出力を参考にします。

最初の2行は、指示を再表示しているようなものです。その下に負荷量（因子パターン）の行列が示されています。以下のように書いてありますね。

Standardized loadings (pattern matrix) based upon correlation matrix

その下が行列なのですが、これまでにSPSSなどで因子分析を行った経験のある人なら違和感があると思います。

```
Call: fa(xb, nfactors = 3, rotate = "promax")  
Standardized loadings (pattern matrix) based  
item ML3 ML1 ML2 h2 u2  
b12 12 0.741 0.109 -0.005 0.644 0.356  
b17 17 0.721 0.121 0.059 0.654 0.346  
h7 7 0.652 0.191 -0.048 0.573 0.427
```

列名を見ると、最初に項目名があり、`item` は（たぶん）列番号、その後各因子がきて、`h2`（本来なら h^2 ）つまり共通性、そして `u2`（本来なら u^2 ）つまり独自性が並んでいます。

因子名は、「**Factor 1** とか **F1** とかでないの？」と思う人もいると思いますが、**R** は因子抽出法の略記（つまり、**ML** とか、**PA** とか、**GLS** とか）に番号を付けたものを因子名にしているようです。

さらに、（これは場合によって違いますが…）1, 2, 3 という順番に並ばないこともあります。今回の例がそうなのですが、3, 1, 2 という並びになっています。これはパターンの値とソートの関係のようです。最も高いパターンを行列の左上にもってきて、それを基準にソートしているようです。ちなみに、ソートをしない場合は1, 2, 3 という順で表示されるので、一度確認してください。

パターン行列の下にも、小さな表が2つ出てきます。上の表の、**SS loadings**が「負荷量の二乗和」、**Proportion Var**が「寄与率」、**Cumulative Var**が「累積寄与率」です。下の表は因子間相関です。もちろん因子間相関は、斜行回転の場合に表示されます。

その下は各種適合性に関する情報です。 χ^2 値とか、各種情報が出ています。ここでは、説明は省きますので（苦笑）調べてみてください。

この因子分析について、ひとつだけ注意喚起を。現在のバージョンでは修正されていますが、少し前のバージョンの時に、**sort=TRUE**を行った際の表示が間違っていることがあったようです。ソートはしてくれるのですが、なぜかu2だけがそのままになるというプログラム上のミスです。共通性は因子で説明される部分、それを1から減じたものがu2、すなわち独自性なので、共通性+独自性=1という等式が成り立ちます。

古いバージョンのまま使っている場合は、少し気をつけておいた方がよい点だと思います。

さて、Rの因子分析の出力では、項目名しか確認できません。これでは項目のまとまりから因子を推測することが難しくなってしまいます。そこで、実際の項目内容を見ながら解釈を進めたいという場合の（できるだけ簡単…だと思ふ）やり方を紹介しておきます。

まずエクセルで新しいファイルを開いておきます。次に、右図のように項目名と項目内容を記していきます。

次に、隣のC列を空欄にしておいて、D1に
=VLOOKUP(C1,\$A\$1:\$B\$20,1,0)と入力します。さらにE1に=VLOOKUP(C1,\$A\$1:\$B\$20,2,0)と入力します。どちらも「#N/A」と表示されますが、この段階では気にせず進みます。この入力したD1、E1の内容を項目b20の位

	A	B
1	b1	親近感のある
2	b2	おしゃれな
3	b3	高級感のある
4	b4	便利な
5	b5	洗練された
6	b6	高価な
7	b7	こだわりがある
8	b8	無機質な
9	b9	使い慣れた
10	b10	かわいい
11	b11	近寄り難い
12	b12	有名な
13	b13	機能的な
14	b14	都会的な
15	b15	安心感のある
16	b16	そそられる
17	b17	安定している
18	b18	リッパな
19	b19	便利な
20	b20	古典的な
21		

	A	B	C	D	E
1	b1	親近感のある		#N/A	#N/A
2	b2	おしゃれな		#N/A	#N/A
3	b3	高級感のある		#N/A	#N/A
4	b4	便利な		#N/A	#N/A
5	b5	洗練された		#N/A	#N/A

置、つまり20行目(D20, E20)までコピーしておきます。

以上でエクセル側の設定は終わりです。

Rに移って、因子分析を実行します。その結果（パターン行列の部分）をコピーし、エクセルのどこか空いているところ（先の項目リストとは別のシートでも構いません）にペーストしてください。ペースト後には、ウィザードを使って、きれいに区切っておきましょう。右図のような感じです。

H	I	J	K	L	M	N	O	P
		item	ML1	ML2	ML3	h2	u2	
	b5	5	0.797	0.047	-0.317	0.737	0.263	
	b14	14	0.753	0.131	-0.327	0.691	0.309	
	b15	15	0.751	-0.089	0.266	0.642	0.358	
	b17	17	0.738	0.195	0.269	0.654	0.346	
	b4	4	0.714	0.191	0.171	0.576	0.424	
	b12	12	0.709	0.253	0.278	0.644	0.356	
	b7	7	0.681	0.27	0.193	0.573	0.427	
	b20	20	-0.627	0.02	0.458	0.604	0.396	
	b16	16	0.592	0.156	0	0.374	0.626	
	b8	8	0.493	-0.203	-0.172	0.313	0.687	
	b2	2	0.488	0.342	0.024	0.356	0.644	
	b18	18	0.442	-0.338	-0.044	0.311	0.689	
	b13	13	0.425	0.167	-0.326	0.314	0.686	
	b10	10	0.407	0.232	-0.009	0.22	0.78	
	b1	1	0.336	0.208	-0.277	0.233	0.767	
	b11	11	0.452	-0.706	0.107	0.714	0.286	
	b3	3	0.375	-0.587	0.101	0.495	0.505	
	b6	6	0.553	-0.572	0.054	0.635	0.365	
	b19	19	0.23	0.181	0.465	0.302	0.698	
	b9	9	0.318	0.182	0.406	0.299	0.701	

そして変数の部分をコピーし、先ほど空けておいたC1からC20の部分にペーストします。すると、DおよびE列に、Cと同様に並び替えられた変数名と項目が表示されます。

これで因子分析の結果順に並び替えられた項目リストができました。後はこの部分をコピーして、必要なところにペーストしてください。もちろん、普通にペーストするのではなく、メニューバーから「形式を選択してペースト」、そして「値」を選んでおきます。

これを作っておけば、後はとても楽になると思います。いろいろと指定を変えて因子分析をし、最も適当なものを探してください。

	A	B	C	D	E
1	b1	親近感のある	b5	b5	洗練された
2	b2	おしゃれな	b14	b14	都会的な
3	b3	高級感のある	b15	b15	安心感のある
4	b4	便利な	b17	b17	安定している
5	b5	洗練された	b4	b4	便利な

●=VLOOKUP(C1,\$A\$1:\$B\$20,1,0)の部分の説明を簡単におきます。関数名が示すように、LOOK-UPするわけですが、カッコ内は「,」で4つに区切られます。その最初が「何を探すか」です。次が「どこを探すか」。3つ目が「どこを探すか」で指定されたエリアの「何列目を結果として表示するか」。4つ目は、とりあえず0にしておけばよい、と考えておいてください。つまり、(C1,\$A\$1:\$B\$20,1,0)は、C1と一致する行を\$A\$1:\$B\$20の範囲から検索して、その範囲の1列目（つまり今回ならA列）に入っているデータを表示しなさいということになります。同様に=VLOOKUP(C1,\$A\$1:\$B\$20,2,0)は、その範囲の2列目（つまりB列）に入っているデータを表示してくれます。

Mac 的コネタ…（かなり無理やり）

統計の本などで軸の回転が説明される時、初期解を2次元に表示して、「これに因子軸を添わせるために回転を…」といった説明を見たことがあるのではないのでしょうか。（Rでもできますが）Macで初期解の3次元の分布を簡単に見ることができます。

やり方は簡単です。まず因子分析で、3因子を抽出し、その際に rotate="none" として、回転前、つまり初期解を求めます。この結果をエクセルに移して整理します。次に、たぶん使ったことがない人が多いと思いますが、「アプリケーション」の「ユーティリティ」に入っている Grapher というプログラムを起動します。

するとどのような新規グラフを作るかを聞いてきますので、3D グラフ、そしてとりあえず Default を選びます。

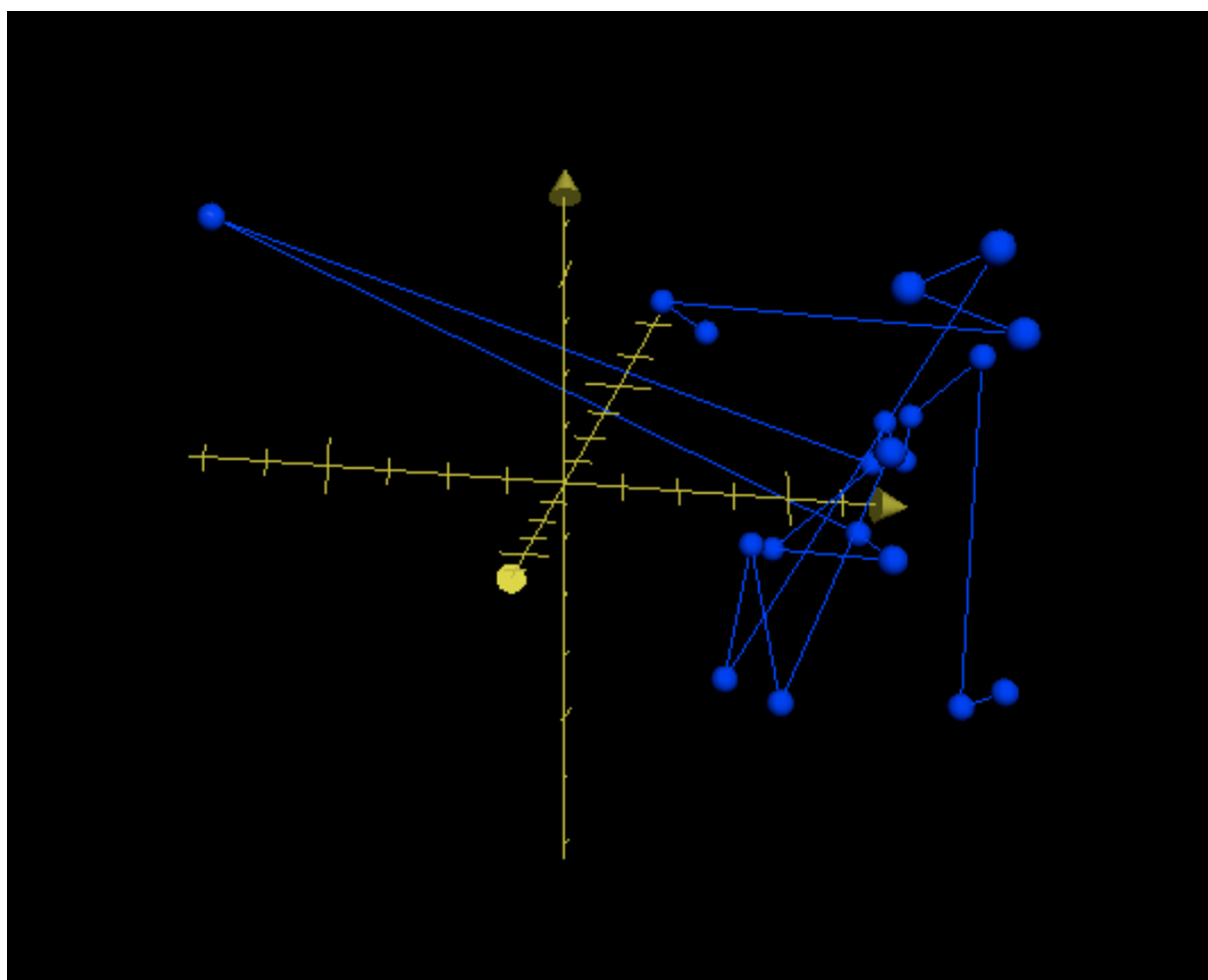
メニューバーの「方程式」から「新規ポイントセット」を選ぶと、4つの点が3D グラフ



内に現れると思います。ここまでくれば、「このポイントセットというデータの部分に、計算しておいた初期解を入れれば、それを3Dにしてくれそう…」という想像が働くのではないのでしょうか。

ポイントは、グラフの上にある「ポイントを編集」というところを変更できます。すでに入っているデータを削除しておいてから、エクセルの初期解の該当部分をコピー、そして先ほど削除した部分にペーストすれば、きれいに数字が入っていきます。そして「OK」とすれば、初期解の3Dプロットが完成。拡大・縮小で見やすい大きさにしてみてください。グラフ内をグリグリすれば、動かして見ることができます。

各データを結ぶ線がちょっと邪魔ですが、3因子抽出の場合だと、結構おもしろいのではないかと思います。



本日はここまでにしますので、いろいろな手法の分析を試してみてください。