

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ことのある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、投入する変数をわりと楽に変えられるからです。

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

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、かなり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施し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

その下が算出された行列なのですが、これまでに SPSS など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。

```
> f1 <- fa(xb, nfactors=3, fm="ml", rotate="promax")
Loading required namespace: GPArotation
> print(f1, sort=TRUE, digit=3)
Factor Analysis using method = ml
Call: fa(r = xb, nfactors = 3, rotate = "promax", fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
      item   ML1   ML3   ML2   h2   u2   com
b20   20 -0.800  0.228 -0.162 0.604 0.396 1.25
b14   14  0.785  0.026  0.070 0.691 0.309 1.02
b5     5  0.763  0.017  0.174 0.737 0.263 1.10
b13   13  0.633 -0.092 -0.080 0.314 0.686 1.08
```

列名を見ると、最初に項目名があり、item は (たぶん) 列番号、その後に各因子がきて、h2 (h²) つまり共通性、そして u2 (u²) つまり独自性が並んでいます。com は複雑性とよばれるもので、ひとつの因子からのみ影響を受けているようであれば1に近くなり、いわゆる「どっちつかず」だと値が大きくなるもの。単純構造に近づけたいような時にはかなり便利な指標だと思います。

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…) ML 1, ML 2, ML 3 という順番に必ずしも並ばないこともあります。上の例がそうなのですが、1, 3, 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。

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とても難しくなってしまいます。現実的には、実際の項目内容と結果の数値を見ながら判断と解釈を進めることが不可避です。Rの因子分析の出力と実際の項目内容を合わせるための（できるだけ簡単…だと思う）やり方を紹介しておきます。エクセルのVLOOKUP関数を利用したやり方です。

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。

	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	操作性のよい
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	使いやすそう
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	りっぱな
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	操作性のよい		#N/A	#N/A
5	b5	洗練された		#N/A	#N/A

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。次図のような感じです。

	item	ML1	ML3	ML2	h2	u2	com	
b20	20	-0.8	0.228	-0.162	0.604	0.396	1.25	
b14	14	0.785	0.026	0.07	0.691	0.309	1.02	
b5	5	0.763	0.017	0.174	0.737	0.263	1.1	
b13	13	0.633	-0.092	-0.08	0.314	0.686	1.08	
b1	1	0.552	-0.06	-0.145	0.233	0.767	1.16	
b16	16	0.364	0.31	0.042	0.374	0.626	1.98	
b10	10	0.312	0.258	-0.102	0.22	0.78	2.17	
b12	12	0.165	0.69	0.027	0.644	0.356	1.12	
b17	17	0.165	0.668	0.096	0.654	0.346	1.17	
b19	19	-0.308	0.657	-0.033	0.302	0.698	1.42	
b9	9	-0.199	0.632	-0.013	0.299	0.701	1.2	
b7	7	0.249	0.597	-0.014	0.573	0.427	1.34	
b4	4	0.257	0.554	0.076	0.576	0.424	1.45	
b15	15	0.06	0.546	0.393	0.642	0.358	1.84	
b2	2	0.362	0.373	-0.182	0.356	0.644	2.44	
b11	11	-0.174	-0.01	0.903	0.714	0.286	1.07	
b6	6	-0.011	0.035	0.79	0.635	0.365	1	
b3	3	-0.158	0.005	0.752	0.495	0.505	1.09	
b18	18	0.133	-0.011	0.494	0.311	0.689	1.15	
b8	8	0.351	-0.064	0.351	0.313	0.687	2.07	

そして変数の部分（上図選択部分）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。

	A	B	C	D	E
1	b1	親近感のある	b20	b20	古典的な
2	b2	おしゃれな	b14	b14	都会的な
3	b3	高級感のある	b5	b5	洗練された
4	b4	操作性のよい	b13	b13	機能的な
5	b5	洗練された	b1	b1	親近感のある

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。

一度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。

●=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。

本日はここまでにします。分析に利用する変数を変えてみて、結果を比べてみてください(Rの出力は、エクセルに移して、きちんと項目と対応させてチェックできるようにしてから比較のこと。ここでの手抜きは厳禁)。因子抽出方法、回転方法が同じでも、ひとつの項目を削除すると、すべての値が変わります。