11 日目：因子分析（1）

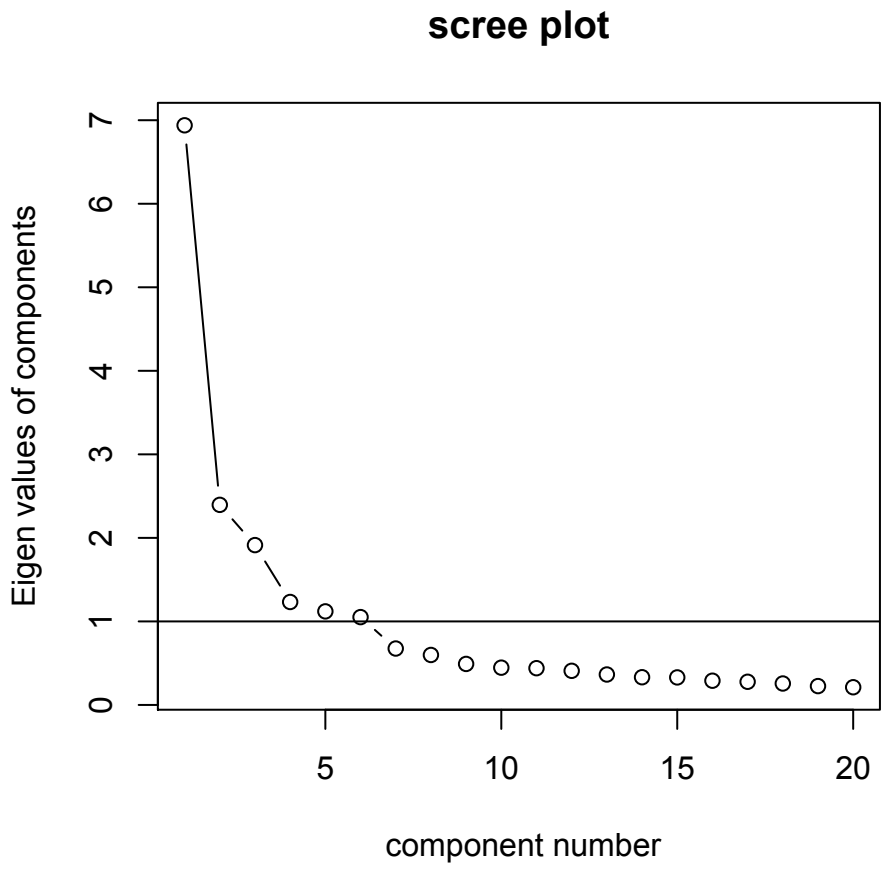
今日は，いよいよ因子分析をやってみます。とはいっても，命令はそれほど難しいものではありません。ただ，手間と時間がかかります…

まず因子分析の第 1 ステップは，固有値を求めて，抽出因子数の目処を立てることです。固有値の推移をみるためには，いくつかのコマンドがありますが，そのうちの２つを紹介しておきます。なお，データは昨日からのものを使います。xb には b1 から b20 までの変数だけを集めてあります。

シンプルな（SPSS と同じ）やり方なら，psych パッケージにある VSS.scree( )を使います。psych パッケージを呼び出しておいて，

```
VSS.scree(xb)
```

これで以下のようなスクリープロットを表示してくれます。なお，VSS.scree(x[5:24])という指定も可能です。



ご丁寧に，固有値１のところに線を引いてくれているので，見やすいと思います。今回のデータならば，最大で６因子です。固有値の推移から判断すると３因子も考えられます。

スクリープロットは目で推移を確認するにはよいのですが，固有値の正確な値を知りたい場合もあるでしょう。このような場合には，直接，固有値を計算させます。

固有値の計算には，eigen( )という関数が準備されています。これは相関マトリックスから計算するようなので，まずは cor で相関マトリックスを準備しておいて，それをeigen( )に渡すという手順でやります。なお eigen( )は，$values と$vectors という２つの結果を表示しますが，固有値は$values の方です。

```
xb.cor <- cor(xb, use="complete.obs")
eigen(xb.cor)
```

この２つを対比すると，確かに固有値がプロットされていることが確認できると思います。

さて，もう１つの方法として，並行分析というものを紹介しておこうと思います。これもpsych パッケージに入っているものです。以下の命令を出してみてください。PC の速度によりますが，ちょっと時間がかかると思います。

```
fa.parallel(xb)
  fa.parallel(x[5:24])  も可
```

場合によって…

```
Loading required package: MASS
```

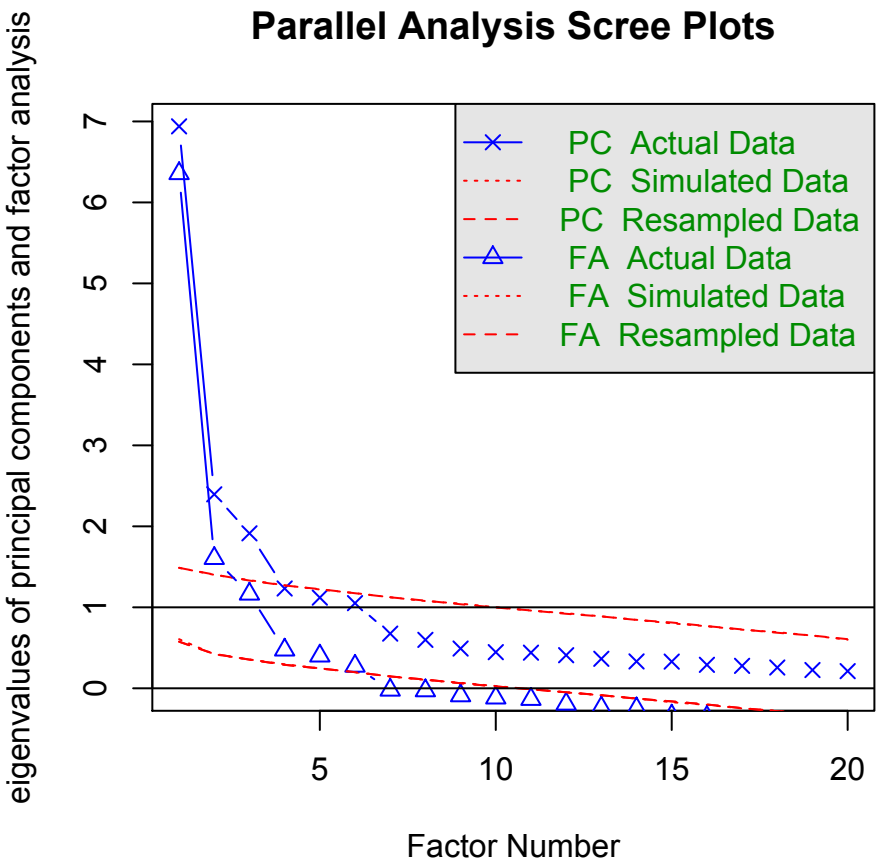
というメッセージが出るかもしれませんが，これは気にしなくてよいです。MASS というパッケージも利用するので，それを組み込みましたということです。どうもRは，パッケージを読み込んでいなければエラーを表示する場合と，自動的に読み込んでくれる場合があるようです。

さらに続けて

```
Parallel analysis suggests that the number of factors =  6  and the
number of components =  3
```

という表示がRコンソールに出てきて，Quartz に新しい図が描かれます。

Rコンソールの表示の方は訳すまでもないでしょうが，「並行分析では，６因子，および３成分が示唆されます」くらいでしょうか。



さてこの平行分析ですが，イメージとしては，分析するデータと同じ変数の数，同じサンプル数の乱数行列を作り，そこから相関行列を導き固有値を求めるという作業を行っています。このランダムな状況から生じた固有値の推移と，分析データの推移を比較し，ランダムなものよりも大きい固有値の個数だけ因子を抽出しようという考え方です。出力される図の中で赤の波線になっている部分が，乱数から作られた固有値の推移です。

`fa.parallel(xb)`は，この平行分析を２つの固有値についてやってくれます。先にやった `VSS.scree(xb)`の結果と見比べると，`PC Actual Data` という「×」マークが`VSS.scree(xb)`の結果と同じことがわかります。論文で一般的に固有値として言及されるものは，この「×」マークの方です。これの推移を乱数で作られた固有値の推移と比較すると，３つ目までは「×」が上に，それ以下は「×」が下に位置しています。ここから，３つの因子をとり出すことが適当という示唆を得ることができます。

もうひとつの「△」の方が因子分析による固有値ですが，私も今一つこの値を理解できていません…。

さて，抽出する因子数に目処が立ったなら，次に因子分析に入ります。

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

たとえばこんな感じです。因子分析は fa()という関数になりますが，この結果を一度何かに保存して，print で出力をコントロールするという方法をとる方がやりやすいでしょう。fa(xb, nfactors=3, fm="pa", rotate="promax")だけでも結果を返してくれますが，因子パターンによる並び替えをしてくれないので，結果を読み取りにくいです。

fa()の中身はというと，まずデータを示すxb（x[5:24]も可），nfactors=で抽出する因子数を，fm=で因子抽出方法を，rotate=で回転の種類を指定します。因子抽出方法としては，"pa"で主因子法，"ml"で再尤法，"gls"で一般化最小二乗法などと指定できます。回転については，プロマックス回転の"promax"，バリマックス回転の"varimax"など様々な指定ができます。回転をしない場合は"none"です。

print の方はだいたい見当がつくと思います。sort=TRUE は並び替えをするかどうか，digit=は負荷量（パターン）の表示桁数をコントロールします。

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。