Factor analysis of the athletics records data

Here we will fit one and two factor models to the athletics records data considered earlier; recall that these data consisted of national records (for men) in 8 running events from 100m to the marathon. A principal component analysis on the correlation matrix (that is, for standardized variables) revealed that the first two principal components had fairly simple interpretations.

The function `factanal` can be used to estimate the loadings and uniquenesses as well as giving an indication of the appropriateness of a given factor model. Maximum likelihood estimation is used and the variables are first standardized to have mean 0 and variance 1; this standardization is not necessary for factor analysis although it does ease interpretation of the loadings.

We will first try a one factor model:

```
> r0 <- factanal(~m100+m200+m400+m800+m1500+m5000+m10000+mara,factors=1)
> r0
Uniquenesses:
  m100   m200   m400   m800  m1500  m5000 m10000   mara
 0.447  0.389  0.242  0.114  0.059  0.104  0.369  0.205
Loadings:
       Factor1
m100   0.743
m200   0.782
m400   0.871
m800   0.941
m1500  0.970
m5000  0.946
m10000 0.794
mara   0.892

               Factor1
SS loadings      6.070
Proportion Var   0.759

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 121.32 on 20 degrees of freedom.
The p-value is 1.63e-16
```

The very small p-value strongly suggests that the one factor model is not adequate. Let's try a two factor model:

```
> r1 <- factanal(~m100+m200+m400+m800+m1500+m5000+m10000+mara,factors=2)
```

```
> r1
Uniquenesses:
  m100   m200   m400   m800  m1500  m5000 m10000   mara
 0.076  0.123  0.147  0.128  0.081  0.037  0.333  0.092
Loadings:
       Factor1 Factor2
m100   0.295   0.915
m200   0.378   0.857
m400   0.552   0.740
m800   0.711   0.605
m1500  0.807   0.517
m5000  0.904   0.382
m10000 0.738   0.350
mara   0.913   0.271
               Factor1 Factor2
SS loadings      3.887   3.095
Proportion Var   0.486   0.387
Cumulative Var   0.486   0.873

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 15.03 on 13 degrees of freedom.
The p-value is 0.306
```

The p-value is much larger here than it is for the one factor model; this indicates that the two factor model fits the data quite well.

The function `factanal` uses the varimax rotation as its default. The loadings for the two factors here are essentially mirror images of each other, the first giving higher loadings to longer distances with the second giving higher loadings to shorter distances.

We will now try other rotations. First of all, we will set `rotation="none"`, which applies no rotation.

```
> r2 <- factanal(~m100+m200+m400+m800+m1500+m5000+m10000+mara,factors=2,
+   rotation="none")
> r2
Uniquenesses:
  m100   m200   m400   m800  m1500  m5000 m10000   mara
 0.076  0.123  0.147  0.128  0.081  0.037  0.333  0.092
Loadings:
       Factor1 Factor2
m100    0.787   0.552
m200    0.818   0.456
```

```
m400    0.887   0.258
m800    0.932
m1500   0.956
m5000   0.952  -0.240
m10000  0.800  -0.165
mara    0.892  -0.334
               Factor1 Factor2
SS loadings      6.198   0.785
Proportion Var   0.775   0.098
Cumulative Var   0.775   0.873

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 15.03 on 13 degrees of freedom.
The p-value is 0.306
```

Note that the loadings here are similar to the loadings for the first two principal components (although the signs and magnitudes are different.

Next we will apply the promax rotation (`rotation="promax"`). This is an oblique rotation (as opposed to an orthogonal rotation) and as a result, the two resulting factors are correlated.

```
> r3 <- factanal(~m100+m200+m400+m800+m1500+m5000+m10000+mara,factors=2,
+  rotation="promax")
> r3
Uniquenesses:
  m100   m200   m400   m800  m1500  m5000 m10000   mara
 0.076  0.123  0.147  0.128  0.081  0.037  0.333  0.092
Loadings:
       Factor1 Factor2
m100   -0.145   1.063
m200            0.935
m400    0.306   0.675
m800    0.601   0.399
m1500   0.783   0.223
m5000   0.994
m10000  0.788
mara    1.072  -0.173
               Factor1 Factor2
SS loadings      3.847   2.700
Proportion Var   0.481   0.337
Cumulative Var   0.481   0.818
```

```
Factor Correlations:
        Factor1 Factor2
Factor1   1.000   0.735
Factor2   0.735   1.000

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 15.03 on 13 degrees of freedom.
The p-value is 0.306
```

These loadings have a similar pattern to the loadings produced by the varimax rotation. Note that because the promax rotation is not an orthogonal rotation, the two factors are correlated.

Note that the overall fit of the two factor model does not depend on the rotation used; the idea behind rotating the loadings is to hopefully find factors (uncorrelated or not) whose loadings are easily interpretable.