R instructions for the 4th seminar

Data set*Satisfaxtion.RData* provides data from questionnaire which aimed at detecting the level of satisfaction in particular areas of life. Sample consisted of 100 randomly choosen adults who answered 10 questions (variables). Higher value of particular variable, higher level of satisfaction.

Analyze the associations among variables. Do exists factors which effects the level of satisfaction in particular areas of life? Is it possible to interpret the factors in a reasonable way?

R Instructions for the problem 1:

•Get familiar with data (correlation matrix,scatterplot matrix,...)

```
!library(DescTools )
!library(ellipse)
!library(car)

x<-cor(Satisfaction,use="pairwise.complete.obs")
pairs(Satisfaction,panel=panel.smooth)
scatterplotMatrix(Satisfaction,smooth=F,diagonal="histogram",col=c(2,1,4))
PlotCorr(x)
plotcorr(x)
```

R Instructions for the problem 2:

•Is FA appropriate method for our data? (=Is correlation matrix significantly different from unit matrix?)

Visually asses graphs provided by correlation plots from previous task. (There exists test, no idea where in R)

R Instructions for the problem 3:

Set the number of factors k.
Firstly the factors will be considered equal with principal components. So find the eigenvalues of correlation matrix of considered data and according to criteria from the third seminar determine the appropriate number of first k principal components.

```
p<-prcomp(x=Satisfaction, center=T,scale.=T)
a<-p$ sdev ^ 2
sum(a[1:2])/10
```

the result=0.7919 expresses percentage of total variance when using first two components
or `summary(p)` - in an output in a "cumulative proportion" line.

R Instructions for the problem 4:

•Create the object in R bearing all essential results of FA for $k = 2$ and $k = 3$.
Set the extraction method to be "maximum likelihood" and rotation method "varimax" je normalizovany.
(We are looking for those rotation methods which lead to loadings approching eighter 0 or +/-1.)

Rather then using `factanal` from the package `stats` use function `fa` from the package `psych`. Both functions performs factor analysis on standardized data (e.g. on correlation matrix). The `factanal` function offers only log likelihood (assuming multivariate normality over the uniquenesses) extraction method, whereas `fa` offers also further extraction methods (=factoring methods) including principal component method.

```
!library(psych)
```

```
f1<-fa(r=x,nfactors=2,rotate="varimax",fm="ml",scores="regression",residuals=T)

f2<-fa(r=x,nfactors=3,rotate="varimax",fm="ml",scores="regression",residuals=T)

f1
```
(`f1$e.values` gives the same result as `a<-p$sdev2`).

R Instructions for the problem 5:

•Express original standardized variables by means of factor variables; interpret the loadings. Do it for $k = 2$.
(e.g.: standardized $AtWork1 = 0.75f_1 + 0.05f_2 + \varepsilon_1$; 0.75 is a correlation of...?)

R Instructions for the problem 6:

•Represent graphically all 10 original variables in a new system of two/three factors. Try to interpret factors (if it gives sense - it is not always reasonable).

```
plot(f1)
plot(f2)
```

R Instructions for the problem 7:

•Interpret communalities. Do it for $k = 2$.
(See the column "h2". E.g. 0.57 in the first line means that 57% out of the total variance of the variable "AtWork1" can by explained by two factors. 0.43 is the unique part of its variance - see colmn "u2".)

R Instructions for the problem 8:

•Asses and interpret the values in residual matrix.

```
f1$residual
```

R Instructions for the problem 9:

•Find the factor score coefficients.

`f1$weights` (e.g. $F_1 = 0.12599\ AtWork1 + \ldots + 0.113355\ General2$)

R Instructions for the problem 10:

•nefakchci prepocitani v nove soustave.