Master in *Statistical Sciences*

Lab 1 *Latent variable models*

Exercise 1
Factor analysis on mental ability score tests

The classic Holzinger and Swineford (1939) dataset consists of mental ability test scores of seventh- and eighth-grade children from two different schools (Pasteur and Grant-White). In the original dataset (available in the MBESS package), there are scores for 26 tests. However, a smaller subset with 9 variables is more widely used in the literature. Perform first an exploratory factor analysis and then a confirmatory factor analysis on this data set

1. Load and display the data set `HolzingerSwineford1939` (library `lavaan`).

   library(lavaan)
   data(HolzingerSwineford1939)
   ?HolzingerSwineford1939
   head(HolzingerSwineford1939)
   data<-HolzingerSwineford1939[,c("x1","x2","x3","x4","x5","x6","x7","x8","x9")]

2. Explore the correlation matrix of the data in order evaluate if a factor model can be fitted.

   matcor<-cor(data)

3. Estimate a factor model with 1,2 and 3 factors using the Maximum Likelihood method and select the best model using the Chi-square test. Use the function `factanal`.

   f1<-factanal(~x1+x2+x3+x4+x5+x6+x7+x8+x9,data=data,1)
   f2<-factanal(~x1+x2+x3+x4+x5+x6+x7+x8+x9,data=data,2)
   f3<-factanal(~x1+x2+x3+x4+x5+x6+x7+x8+x9,data=data,3)
   Chisq<-round(c(f1$STATISTIC,f2$STATISTIC,f3$STATISTIC),3)
   df<-c(f1$dof,f2$dof,f3$dof)
   ***p*-value<-round(c(f1$PVAL,f2$PVAL,f3$PVAL),4)**

4. On the basis of the matrix of the factor loadings interpret the factors individuated

   loadings(f3)
   print(f3,cutoff=0.3)

5. Compute the communalities and comment them. Which percentage of the variance of the model is explained by the three-factor model?

   comm<-rowSums(loadings(f3)^2) or comm<1-f3$uniquenesses
   percVar<-sum(comm)/9

5. Compute the reproduced correlation matrix and the discrepancy between the observed and reproduced correlation.

   repcorr<-loadings(f3)%*%t(loadings(f3))
   round(matcor-repcorr,3)

6. Apply different orthogonal and oblique rotations and interpret the solutions obtained

   library(GPArotation)
   Varimax(loadings(f3))
   quartimax(loadings(f3))
   oblimin(loadings(f3))

7. Compute the factor scores using the Bartlett and the Thompson method

   fB<-factanal(~x1+x2+x3+x4+x5+x6+x7+x8+x9,data=data,3,scores="Bartlett")$scores
   fT<-factanal(~x1+x2+x3+x4+x5+x6+x7+x8+x9,data=data,3, scores="regression")$scores

8. On the basis of the previous analysis perform a confirmatory factor analysis using the function `cfa` in lavaan.

   HS.model <- ' visual =~ x1 + x2 + x3
   textual =~ x4 + x5 + x6
   speed =~ x7 + x8 + x9 '

   fit <- cfa(HS.model, data=data)
   summary(fit,fit.measures=TRUE)

   fit <- cfa(HS.model, data = HolzingerSwineford1939, std.lv = TRUE)

   fit.HS.ortho <- cfa(HS.model, data = HolzingerSwineford1939, orthogonal = TRUE)

Exercise 2
Social mobility in UK

The data are based on information provided by 713 male or female married respondents to a survey carried out in 1949. The variables relate to the respondent, his or his spouse, father, father-in-law, and firstborn son. The file socmob.txt contains the full correlation matrix.

The 10 variables are coded as follows:
$X1$=Husband's father's occupational status
$X2$=Wife's father's occupational status
$X3$=Husband's further education
$X4$=Husband's qualifications
$X5$=Husband's occupational status
$X6$=Wife's further education
$X7$=Wife's qualifications
$X8$=Firstborn's further education
$X9$=Firstborn's qualifications
$X10$=Firstborn's occupational status

Following the steps of the analysis of the previous example perform a factor analysis.

1. Load the correlation matrix `socmob.txt`.
2. Perform an exploratory factor analysis.
3. Comment the obtained results.