Lab 3 ***Latent class analysis***

Exercise 2
Women's mobility data

The data are from the Bangladesh Fertility Survey of 1989. The questionnaire contains a number of items that indicate whether a woman living in rural Bangladesh could engage in various activities alone. Following the steps of exercise perform a latent class analysis in order to identify groups of women with similar patterns of mobility.

1. Load the data set Mobility (library ltm).
   ?Mobility

2. Recode the data set properly in order to use the function poLCA.
   names(Mobility)<-c("Item1","Item2","Item3","Item4", "Item5","Item6","Item7","Item8")
   Mobility[Mobility==1]<-2
   Mobility[Mobility==0]<-1

3. Use the function poLCA to fit the models with 2, 3 e 4 latent classes to the data.
   library(poLCA)
   formula<-cbind(Item1,Item2,Item3,Item4,Item5,Item6,Item7,Item8)~ 1
   m.2<-poLCA(formula,Mobility,nclass=2,nrep=10,verbose=FALSE)
   m.3<-poLCA(formula,Mobility,nclass=3,nrep=10,verbose=FALSE)
   m.4<-poLCA(formula,Mobility,nclass=4,nrep=10,verbose=FALSE)

4. Compare the estimates of the expected frequencies for the response pattern obtained with the three models.
   freq.estim<-data.frame(m.2$predcell[1:10],m.3$predcell[10],m.4$predcell[10])
   freq.estim

5. Compare the three models considering the following measures: log-likelihood, number of free parameters, likelihood ratio test, Pearson chi-square test (and associated p-value), AIC, BIC.
   K<-c("2", "3", "4")
   llik<-c(m.2$llik,m.3$llik,m.4$llik)
   npar<-c(m.2$npar,m.3$npar,m.4$npar)
   Gsq<-round(c(m.2$Gsq,m.3$Gsq,m.4$Gsq),3)
   Chisq<-round(c(m.2$Chisq,m.3$Chisq,m.4$Chisq),3)
   df<-c(m.2$resid.df,m.3$resid.df,m.4$resid.df)
   pvalue<-round(1-pchisq(Chisq,df),4)

AIC<-round(c(m.2$aic,m.3$aic,m.4$aic),3)
BIC<-round(c(m.2$bic,m.3$bic,m.4$bic),3)
summary<-data.frame(K,llik,npar,Gsq,Chisq,df,pvalue,AIC,BIC)

6. Which model shows the best fit to the observed data?

7. Display the parameter estimates of the four-class model and compare this solution with those obtained with your colleagues. Which problem arises?
round(m.4$P,4)
lapply(m.4$probs,round,4)

8. Use the function poLCA.reorder to solve the indeterminacy of the ranking of the latent classes.
probs.start.m4<-m.4$probs.start
new.probs.start.m4<-poLCA.reorder(probs.start.m4,order(m.4$P))
m.4.ord<-poLCA(formula,Mobility,nclass=4,probs.start=new.probs.start.m4 ,verbose=FALSE)
round(m.4.ord$P,4)
lapply(m.4.ord$probs,round,4)

9. Interpret the four latent classes.

10. Display the posterior probability estimates of the response pattern (1,1,1,1,1,1,1,1) and (2,2,2,2,2,2,2,2) for the four latent classes. In which classes the samples of women corresponding to these response patterns are allocated?
a<-as.numeric(row.names(Mobility[which(Mobility$Item1==1&Mobility$Item2==1&Mobility$Item3==1&Mobility$Item4==1&Mobility$Item5==1&Mobility$Item6==1&Mobility$Item7==1&Mobility$Item8==1),]))
post.1<-m.4.ord$posterior[a[1],]
round(post.1,5)
b<-as.numeric(row.names(Mobility[which(Mobility$Item1==2&Mobility$Item2==2&Mobility$Item3==2&Mobility$Item4==2&Mobility$Item5==2&Mobility$Item6==2&Mobility$Item7==2&Mobility$Item8==2),]))
post.2<-m.4.ord$posterior[b[1],]
round(post.2,5)

11. Display the latent class in which the samples of women are allocated according to highest posterior probability.
m.4.ord$predclass
table(m.4.ord$predclass)

12. Select the samples of women allocated to the class with less mobility and evaluate the corresponding response patterns.
    cl3<-Mobility[m.4.ord$predclass==3,]
    unique(cl3)

13. Compute the predicted cell probability from the latent class model of the response pattern (2,1,2,1,2,1,2,1), that is not observed in the dataset, and allocate it according to the highest posterior probability.
    poLCA.predcell(m.4.ord,c(2,1,2,1,2,1,2,1))
    poLCA.posterior(m.4.ord,y=c(2,1,2,1,2,1,2,1))