Master in *Statistical Sciences*

Lab 2 *Latent variable models*

Exercise 1
Attitude towards abortion

The data contain responses given by 410 individuals to four out of seven items concerning attitude to abortion. A small number of individual (379) did not answer to some of the questions and this data set contains only the complete cases.

1. Load and display the *dataframe* `Abortion` (library `ltm`).

2. Use the function `descript` to summarize the principal characteristics of the data set dataset. In particular evaluate:
- which one among the 4 items shows the lowest degree of agreement among the sampled individuals; - what is the distribution of the *total score* in the sample; - if there are problems in some pairs of items ;

library(ltm)
data(Abortion)
?Abortion
dim(Abortion)
head(Abortion)
?descript
dsc <- descript(Abortion)
dsc$perc
dsc$items
dsc$pw.ass

2. Use the function `ltm` to fit **the 2 IRT model** to the data. Comment the results. In particular describe :

2.1 Which item shows the highest positive attitude towards the abortion by the interviewed individuals

2.2 Which item more discriminate among individuals in favour and against the abortion

2.3 Which item less discriminate among individuals in favour and against the abortion

2.4 The interpretation of the latent variable

?ltm
m1<-ltm(Abortion ~ z1)

m1.rip<-ltm(Abortion~ z1,IRT.param=FALSE)
summary(m1)
summary(m1.rip)

2.5 Verify the relation between the parameter estimates in the two parameterizations.
-summary(m1)$coefficients[1,1]*summary(m1)$coefficients[5,1]
summary(m1.rip)$coefficients[1,1]

4.3 Compute the standardized discriminant parameters and the probability of positive response of the median individual
alpha<-m1.rip$coeff[,2]
stalpha<-alpha/sqrt(1+alpha^2)
coef(m1.rip,prob=TRUE,order=TRUE)

6. Represent the characteristic curves of the four `items` for the estimated model. Comment the results.

plot(m1, legend = TRUE, cx = "bottomright", xlab="Attitude toward abortion", lwd = 3, cex.main = 1.5, cex.lab = 1.3, cex = 1.1)

7. Based on the obtained results evaluate if the model has a good fit to the data.

E<-fitted(m1)[,5]
O<- m1$patterns$obs
cbind(m1$pattern$X,m1$pattern$obs,E)
Chisq<-sum((E-O)^2/E)
DOF<-14-2*4-1
pvalueC<-1-pchisq(Chisq,DOF)
LR<-2*sum(O*log**(O/E))**
pvalueLR<-1-pchisq(LR,DOF)

8. Use the function `margins` to compute the values of the two-way and three-way chi-square residuals and evaluate if the model has a local goodness of fit to the data.

?margins
margins(m1)
margins(m1,type="three-way",nprint=2)
margins(m1,type="three-way",nprint=3)

9. Use the function `factor.scores` to estimate the values of the latent variable for each individual.

?factor.scores
fs<-factor.scores(m1,method="EAP")
fs
Comp<-factor.scores(m1,method="Component")
Comp

```
resp.pattern<-fs$score.dat[,1:4]
total.score<-apply(resp.pattern,1,sum)
total.score
Cp<- Comp$score.dat[,7] tab<-cbind(fs$score.dat,Cp,total.score)
round(tab,3)
round(tab[order(total.score),],3)
```

10. Estimate the abilities of the *response pattern* that do not occur in the dataset.

```
fitted(m1,resp.pattern=rbind(c(1,0,0,1),c(1,0,1,0)))
```

Exercise 2
Law School Admission Test

The LSAT is a classical example in educational testing for measuring ability traits. This test was designed to measure a single latent ability scale. This LSAT example is a part of a data set given in Bock and Lieberman (1970).

1. Load and display the *dataframe* `LSAT` (library `ltm`).

2. Use the function `descript` to summarize the principal characteristics of the data set. In particular evaluate:
- which one among the 5 items is the most difficult among the candidates;
- what is the distribution of the *total score* in the sample;
- if there are problems in some pairs of items `item` ;
- which item has the highest/smallest proportion of correct responses given the same value of the *total score*.

dsc <- descript(LSAT)
dsc$perc
dsc$items
dsc$pw.ass
plot(dsc)

3. Use the function `rasch` to fit the **constrained Rasch model** (same discrimination parameter equal to 1 for all the **items).**

?rasch
m1<-rasch(LSAT, IRT.param=TRUE,constraint = cbind(ncol(LSAT) + 1, 1))
m1.rip<-rasch(LSAT, IRT.param=FALSE,constraint = cbind(ncol(LSAT) + 1, 1))

3.1 Display and comment the parameter estimates in the latent trait parameterization.

summary(m1.rip)

3.2 Compute and comment the estimates of the correct probability of the five items by the median individual

coef(m1.rip,prob=TRUE,order=TRUE)

3.3 Use the function `GoF.rash` to estimate the *bootstrap* $p-value$ of the X^2 statistic and establish if the model has an overall goodness of fit to the data.

pval.boot<-GoF.rasch(m1,B=199)
pval.boot$Tobs
pval.boot

3.4 Use the function `margins` to compute the values of the two-way and three-way chi-square residuals and evaluate if the model has a local goodness of fit to the data.

margins(m1)
margins(m1,type="three-way",nprint=2)
margins(m1,type="three-way",nprint=3)

4. Use the function `rasch` to fit **unconstrained Rasch model** ($a \neq 1$).

m2<-rasch(LSAT)
m2.rip<-rasch(LSAT,IRT.param=FALSE)

4.1 Compare the new estimates with the previous ones: what is the effect of removal the constrain on the parameters?

summary(m2.rip)

4.2 Verify the relation between the parameter estimates in the two parameterizations.

summary(m2)$coefficients[1,1]*summary(m2)$coefficients[6,1]
summary(m2.rip)$coefficients[1,1]

4.3 Establish if the discrimination parameter can be considered not equal to 1.

anova(m1,m2)

4.4 Verify if the model has problems of local of fit (two and three way margins).

margins(m2)
margins(m2,type="three-way",nprint=2)

4.5 Use the AIC e BIC criteria for the selection between the two Rasch models.

5. Use the function `ltm` to fit **the 2 IRT model** to the data.

m3<-ltm(LSAT ~z1)
m3.rip<-ltm(LSAT~z1,IRT.param=FALSE)

5.1 Compare the estimates of the discrimination parameters of the items.

summary(m3.rip)

5.2 Evaluate if there is a significant difference between this model and the previous one in terms of goodness of fit.

anova(m2,m3)

6. Represent the characteristic curves of the five items `item` for the selected model. Comment the results.

plot(m2, legend = TRUE, cx = "bottomright", lwd = 3, cex.main = 1.5, cex.lab = 1.3, cex = 1.1)

7. Use the function `factor.scores` to estimate the candidates of the Law School according to the selected model.

```
fs<-factor.scores(m2,method="EAP")
fs
plot(fs$score.dat$z1)
resp.pattern<-fs$score.dat[,1:5]
total.score<-apply(resp.pattern,1,sum)
total.score
round(fs$score.dat[order(total.score),],3)
```

8. Estimate the abilities of the *response pattern* that do not occur in the dataset.

```
factor.scores(m2,resp.pattern=rbind(c(0,1,1,0,0),c(0,1,0,1,0)))
```