Introduction to Latent Class Analysis with Applications

Mariano Porcu

and

Francesca Giambona

University of Cagliari

Supplementary files
Appendix A

LCA in SAS

In order to perform a LCA with SAS, you have to download the PROC LCA and PROC LTA (Latent Trait Analysis) procedures available for free for registered members at The Methodology Center. The Methodology Center is an interdisciplinary research center within the College of Health and Human Development at the University Park campus of Penn State.

Go to the URL http://methodology.psu.edu/, then click on the tab "free software" and then click on the PROC LCA & PROC LTA link in the page and follow the detailed instructions for the installation. Detailed information on the procedure and all the related options can be found on the very detailed PROC LCA & PROC LTA user's guide downloadable from The Methodology Center website.

To fit the LCA with the data provided, you have to upload the database in SAS. To do that, go in the Explorer tab of the SAS graphical interface, double click on your working directory and then File > Import. Next, we select the data source in the list provided clicking on "Tab Delimited File (*.txt)". Select the proper path (clicking on the button "Browse") and choose "lca_data.txt". Then, choose the SAS destination (your working library) and then give a name to the database (i.e., the "member" of the library).

Now, we can fit a latent class model by calling the PROC LCA procedure. To do that we have to enter the following code:

```
proc lca data=lcasi.lca_data; /*specify the data to be modeled*/
 nclass 3;                    /*specify the # of the latent classes (≥1)*/
 items QUIETPLACE             /*enter the list of manifest variable names*/
       COMPUTER
       DESK
       WEB
       ENCY
       BOOKS
       JOYREAD
       SPORT
       CULTACT
       LANG;
 categories 2                 /*number of categories of the manifest variables*/
            2                 /*above listed (follow the same order as above)*/
            2
            2
            2
```

```
            3
            3
            2
            2
            2;
 rho prior = 1;              /*see Lanza et al., 2014*/
 seed 1;                     /*random starting values for the ρ parameters
                             can be*/
                             /*generated in PROC LCA by specifying a positive*/
                             /*integer value in the SEED statement. An
                             integer*/
                             /*seed to generate the random values allows
                             the user*/
                             /*to replicate the analysis*/
run;
```

You can fit a LCA with a different number of classes by simply entering an integer number to specify the desired number of latent classes in “nclass”.

LCA in R

To run LCA with R, you need to download the poLCA package from the CRAN archive. Once you have launched R in your machine point to the menu “Packages,” next click on “Install Package”. If you have not set it before, the system will ask you to select a repository (the “mirror”) from where the package will be uploaded; select the one nearest to you. Now, scroll the list (it is in alphabetical order) and select poLCA. Now, in the menu “Packages” click on “Load Package” and select poLCA scrolling the list that opens. Alternatively, at the URL http://cran.r-project.org/ click on the link “Packages” in the left hand menu. In the next page click on “Table of available packages sorted by name”, click on poLCA and download the package source zip file. Detailed explanations on poLCA package can be find in the available pdf manual downloadable from CRAN.

First, you have to upload the dataset. Then, enter the following code:

```
lcadata <- read.table("C://SPECIFY_YOUR_PATH/lca_data.txt", sep="\t",
header=TRUE)
```

Next, type:

```
attach(lcadata)
dim(lcadata)
```

Set all the columns (except the first one that contains the record ID) in the numeric format:

```
for(i in 2:ncol(lcadata)){lcadata[,i]<-as.numeric(lcadata[,i])}
```

Check if the above selected columns are in the numeric format:

```
apply(lcadata[,2:20],2,mean, na.rm=TRUE)
```

Now, we perform the LCA by fitting models with an increasing number of classes. The number of classes is set with the argument "nclass= ". NOTE: all the models fitted contain no covariates.

```
f.lcanocov <-
cbind(QUIETPLACE,COMPUTER,DESK,WEB,ENCY,BOOKS,JOYREAD,SPORT,CULTACT,LANG)~1

f.lc1nocov <- poLCA(f.lcanocov,lcadata,nclass=1,nrep=9,na.rm=FALSE)
f.lc2nocov <- poLCA(f.lcanocov,lcadata,nclass=2,nrep=9,na.rm=FALSE)
f.lc3nocov <- poLCA(f.lcanocov,lcadata,nclass=3,nrep=9,na.rm=FALSE)
f.lc4nocov <- poLCA(f.lcanocov,lcadata,nclass=4,nrep=9,na.rm=FALSE)
```

Now we draw the scree plot entering the values of the BIC statistics as computed in the previous step:

```
LCX <- c(1,2,3,4,5)
BIC <- c(23470.10,23046.16,22906.53,22933.15)
```

```
plot (LCX, BIC, xlab="No. of Latent Classes", ylab="BIC", type="b", col=2)
```

With poLCA package, it is also easy to plot the estimated item response probabilities for the selected 3 classes model:

```
f.lc3nocov <-
cbind(QUIETPLACE,COMPUTER,DESK,WEB,ENCY,BOOKS,JOYREAD,SPORT,CULTACT,LANG)~1

lc3nocov <- poLCA(f.lc3nocov,lcadata,nclass=3,nrep=5,na.rm=FALSE, graphs=T)
```