

Introduction to Latent Class Analysis

Dr Oliver Perra: o.perra@qub.ac.uk

We are going to use Mplus to conduct some analyses in a dataset:

- `ess_ex3.dat`

In what follows, I will describe the Dataset we will be using for this practical.

Two sets of exercises are then provided. Solutions are available in separate files.

The Dataset

This is another extract of variables from the European Social Survey (Round 7), which have been transformed before analyses. As well as the variables concerning influencing politics, which I have used in previous exercises, there are additional covariates and distal outcomes that I will use in the following exercises.

To recap, the indicators used in estimating the latent class are:

rspspsgv : “Political system allows people to have a say in what government does”
ractrolg : “Able to take active role in political group”
rspppi : “Political system allows people to have influence on politics”
rcptppol : “Confident in own ability to participate in politics”
rptcpplt : “Politicians care what people think”
retapapl : “Easy to take part in politics”

All these variables have been recoded into 3 categories (**0; 1; 2**), with higher scores representing higher level of agreement with the statement.

The dataset includes only respondents from 6 countries: Austria, Belgium, Switzerland, Spain, and France. Overall, there are 8,938 respondents in the dataset

Other background variables in the dataset set are:

ess7id : A numerical ID.
cou : A variable indicating the respondent’s country (1 = Austria [AT]; 2= Belgium [BE]; 3=Switzerland [CH]; 8=Spain [ES]; 10=France [FR]).
nuts1en : An ID to identify different NUTS1 areas.
nuts2en : An ID to identify different NUTS2 areas.

- The **covariates** in the dataset are:

male : a categorical variable where 1=respondents who are identified as males , and 0=respondents who were not identified as males
agec : a continuous variable representing respondents' age , centred at 30 years.
dcou1 : A dummy-coded variable where 1= respondents from AUSTRIA , and 0=respondents from other countries
dcou2 : A dummy-coded variable where 1= respondents from BELGIUM , and 0=respondents from other countries
dcou3 : A dummy-coded variable where 1= respondents from SWITZERLAND , and 0=respondents from other countries
dcou4 : A dummy-coded variable where 1= respondents from SPAIN , and 0=respondents from other countries
dcou5 : A dummy-coded variable where 1= respondents from FRANCE , and 0=respondents from other countries
dedu1 : A dummy-coded variable where 1=respondents with no educational attainment or educational attainment below General ISCED 3 >= 2 years , and 0 =respondents with higher educational achievements
dedu2 : A dummy-coded variable where 1=respondents with educational attainment equal or above General ISCED 3 >= 2 years, but below ISCED 5A (short, intermediate/academic/general tertiary below bachelor) , and 0 =all other respondents
dedu3 : A dummy-coded variable where 1=respondents with educational attainment equal or above ISCED 5A (short, intermediate/academic/general tertiary below bachelor) , and 0 =all other respondents.

To summarise, the **dcou...** variables represent dummy-codings of the “**cou**” (country) variable, which can be used as covariates in the analyses. Similarly, the **dedu...** variables are dummy coding of the variable educational attainment, where **dedu1** (for dummy-educational attainment) represents individuals with no or lower levels of educational attainment, and **dedu3** represents individuals with tertiary degrees or higher level of educational achievement.

- The **distal outcome** in the dataset is:

happy : This variable indicates how happy the respondent reported being in a scale from 0 (“Extremely unhappy”) to 10 (“Extremely happy”).

The dataset represents cross-sectional information, therefore happiness cannot really be considered a “distal” outcome, but I will use this variable *as if it were* a distal outcome just for the sake of example.

Exercise #3

Multiple Pseudo-Class Draws

TASKS:

1. Use the “ess_ex3_0.inp” as your start point to test a model where the 4 latent classes with indicators **rpsppsgv ractrolg rpsppiyl rcpttpol rptcpplt retapapl** are regressed on covariates for gender, age, country, and educational attainment (HINT: dummy variable), using the Mplus options to invoke a multiple pseudo-class draw approach.
2. Using the results from task 1, report the odds ratios of being in the “Optimist” latent class rather than the “Sceptical” one for people with the lower level of educational level (**edu1**) compared to those with the highest educational level (**edu3**)
3. Use Mplus options to invoke Multiple Pseudo-Class Draws to test differences in the average happiness (variable: **happy**) across the 4 latent classes.
4. Using the results from task 3, which latent class display the highest average happiness, and is this average significantly different from that of participants in the other classes?

Exercise #4

Three-Step Approach

TASKS:

1. Use the “ess_ex3_0.inp” as your start point to estimate a model with 4 latent classes using indicators **rpsppsgv ractrolg rpsppiyl rcpttpol rptcpplt retapapl**, and create a datafile that includes the most likely latent class as well as the variables **happy agec male dcou1 dcou2 dcou3 dcou4 dcou5 dedu1 dedu2 dedu3**.
2. Use the OUTPUT created in Task 1 to specify a new model where estimation of 4 latent classes is fixed at the measurement parameters obtained in Task 1, and latent class affiliation is regressed on covariates for gender, age, country, and educational attainment.
3. Check the solution of Task 2 above. Do the latent classes estimated correspond in number to those estimated in Task 1?
4. Check the solution of Task 2 above to report odds ratios (and 95% Confidence Intervals) of being in latent class “Optimist” rather than “Sceptical” across countries.
5. Run a new model as in Task 2 above, but also include the distal outcome variable **happy**:
 - a. Model the regression of **happy** on the other covariates in the model;
 - b. Estimate the mean and variance of **happy** across the 4 classes;
 - c. Test whether average **happy** is similar in latent class “Optimist” and “Sceptical”.