

Solutions to 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 rpsppi1 rcpttpol rptcpplt retapap1** are regressed on covariates for gender, age, country, and educational attainment (HINT: dummy variables), using the Mplus options to invoke a multiple pseudo-class draw approach.

ANSWER: Mplus allows to include covariates as predictors of latent class affiliation using specific language in option **AUXILIARY** of command **VARIABLE=**. An example is provided in INPUT file “ess_ex3_1.inp”.

```
Data:
  File is ess_ex3.dat ;
Variable:
  Names are
    essround idno polintr happy rpsppsgv ractrolg rpsppi1 rcpttpol rptcpplt
    retapap1 cou ess7id nutslen nuts2en nuts3en agec edu male dcou1 dcou2
    dcou3 dcou4 dcou5 dedu1 dedu2 dedu3;
  Missing are all (-999) ;
  Usevariables are
    rpsppsgv ractrolg rpsppi1 rcpttpol rptcpplt retapap1;

CATEGORICAL= rpsppsgv ractrolg rpsppi1 rcpttpol rptcpplt retapap1;

AUXILIARY= male(R) dcou1-dcou4(R) agec(R) dedu1-dedu2(R);

Classes=class(4);
```

The line:

```
AUXILIARY= male (R) dcou1-dcou4 (R) agec(R) dedu1-dedu2(R)
```

is ensuring that the covariates listed are considered and tested as predictors of latent class membership using an Mplus built-in option for invoking multiple pseudo-class draws.

It is important to remember that variables like “Gender” and “Country” are nominal, while “Educational Attainment” is an ordered categorical variable: when we regress latent class membership on these covariates, we have to choose a category of these variables as a reference category.

For example, in the case of “Gender”, I had created a dummy-variable **male** that represents whether participants are reported as being male or other. Including **male** as a covariate means that the regressions will compare how the probability of being in one latent class differ between males and others.

As for country, I chose to use France as the reference category: the dummy variable that represented France in the dataset is **dcou5**: by excluding this variable from the **USEVAR=** and the **AUXILIARY=** options, I am ensuring that the coefficients associated with the other dummy-variables

(**dcou1** to **dcou4**) represent changes in latent class probabilities between each of the latter variables and the reference category **dcou5** (France).

In the same manner, I chose to exclude one of the dummy variables that represent different levels of Educational Attainment: since I have excluded **dedu3**, representing the highest level of educational attainment, these individuals will be the reference category in the comparisons carried out by the analyses.

When using continuous covariates, it is often advisable to centre or standardise them. I had centred the age variable to age=30 years when preparing the dataset, creating variable **agec**. However, Mplus provides options to manipulate, centre, and standardise variables with the command

DEFINE:

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)

ANSWER: The results are in OUTPUT file “ess_ex3_1.out”. In the solution I obtained, latent class 3 represents the “Optimist” class, while latent class 4 represents the “Sceptical” class. Note that the order in which the classes appear may change when you run your analyses (due to effects of the random starts).

By default, Mplus use the last latent class as the reference category in multinomial logistic regressions. Since the last class in the solution I obtained is the “Sceptical” class, the results Mplus reports indicate changes in the probability of being in one class compared to the “Sceptical” class, taken as a reference (since it happens to be the last one).

In the Output you will find these results:

TESTS OF CATEGORICAL LATENT VARIABLE MULTINOMIAL LOGISTIC REGRESSIONS USING
POSTERIOR PROBABILITY-BASED MULTIPLE IMPUTATIONS (PSEUDO-CLASS DRAWS)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
CLASS#1 ON				
MALE	0.523	0.067	7.751	0.000
DCOU1	0.139	0.103	1.348	0.178
DCOU2	-0.349	0.109	-3.194	0.001
DCOU3	0.289	0.141	2.054	0.040
DCOU4	0.150	0.097	1.545	0.122
AGEC	-0.013	0.002	-7.206	0.000
DEDU1	-1.427	0.101	-14.165	0.000
DEDU2	-0.844	0.088	-9.582	0.000
CLASS#2 ON				
MALE	0.641	0.075	8.543	0.000
DCOU1	0.430	0.125	3.455	0.001
DCOU2	0.237	0.123	1.932	0.053
DCOU3	2.403	0.127	18.873	0.000
DCOU4	-0.192	0.132	-1.459	0.145
AGEC	-0.014	0.002	-7.023	0.000
DEDU1	-1.636	0.102	-16.035	0.000
DEDU2	-1.418	0.093	-15.271	0.000
CLASS#3 ON				
MALE	0.137	0.063	2.166	0.030
DCOU1	-0.106	0.097	-1.091	0.275
DCOU2	0.217	0.093	2.341	0.019
DCOU3	1.203	0.115	10.495	0.000
DCOU4	-0.485	0.099	-4.904	0.000
AGEC	-0.014	0.002	-8.286	0.000
DEDU1	-0.690	0.091	-7.557	0.000
DEDU2	-0.472	0.086	-5.494	0.000

Which report the coefficients representing changes in the probability of being in latent class 1 rather than latent class 4 (the reference category), changes in the probability of being in latent class 2 rather than latent class 4, etc.

Since latent class 2 is the “Optimist” class and latent class 4 (which also happens to be the reference class) is the “Sceptical” class, I am focusing on these results:

CLASS#2 ON				
MALE	0.641	0.075	8.543	0.000
DCOU1	0.430	0.125	3.455	0.001
DCOU2	0.237	0.123	1.932	0.053
DCOU3	2.403	0.127	18.873	0.000
DCOU4	-0.192	0.132	-1.459	0.145
AGEC	-0.014	0.002	-7.023	0.000
→ DEDU1	→ -1.636	0.102	-16.035	0.000
DEDU2	-1.418	0.093	-15.271	0.000

I have omitted the dummy variable **dedu3**, representing individuals with the highest educational attainment, so the coefficient -1.636 represents changes in probability of being in latent class 2 (“Optimist”) rather than class 4 (“Sceptical”) for participants with no or lower level of educational attainment (**dedu1**) compared to participants with the highest level of educational attainment (**dedu3**, the reference category).

The coefficient -1.636 represents the logit, so by exponentiating it, we obtain odds ratios = 0.195.

In conclusion: Compared to those with the highest level of educational attainment (**dedu3**), those with the lowest level of educational attainment display a 80% (OR = 0.195) reduction in the odds of being in the “Optimist” class (class 2) rather than the “Sceptical” class (class 4). Note that the p value of this comparison is $p < .001$.

3. Use Mplus options to invoke Multiple Pseudo-Class Draws to test differences in the average happiness (variable: happy) across the 4 latent classes.

ANSWER: The solution is provided in INPUT file “ess_ex3_2.inp”, and it involves the combination of option **AUXILIARY=** and other characters in the **VARIABLE:** command.

```

Variable:
  Names are
    essround idno polintr happy rpsppsgv ractrolg rpsppipl rcptppol rptcpplt
    retapapl cou ess7id nutslen nuts2en nuts3en agec edu male dcoul dcou2
    dcou3 dcou4 dcou5 dedu1 dedu2 dedu3;
  Missing are all (-999) ;
  Usevariables are
    rpsppsgv ractrolg rpsppipl rcptppol rptcpplt retapapl;

CATEGORICAL= rpsppsgv ractrolg rpsppipl rcptppol rptcpplt retapapl;

AUXILIARY= happy(E);

Classes=class(4);

```

The line:

AUXILIARY: happy(E);

is invoking a test of the null hypothesis of equal means in variable **happy** across the latent classes estimated using posterior probability-based multiple imputations (pseudo-class draws).

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?

ANSWER: The results are in OUTPUT file “ess_ex3_2.out”. In particular, here:

EQUALITY TESTS OF MEANS ACROSS CLASSES USING POSTERIOR PROBABILITY-BASED
MULTIPLE IMPUTATIONS WITH 3 DEGREE(S) OF FREEDOM FOR THE OVERALL TEST AND
1 DEGREE OF FREEDOM FOR THE PAIRWISE TESTS

HAPPY

	Mean	S.E.		Mean	S.E.
Class 1	7.490	0.046	Class 2	8.100	0.040
Class 3	7.623	0.036	Class 4	7.237	0.035
	Chi-Square	P-Value		Chi-Square	P-Value
Overall test	247.357	0.000	Class 1 vs. 2	98.723	0.000
Class 1 vs. 3	4.809	0.028	Class 1 vs. 4	18.339	0.000
Class 2 vs. 3	73.313	0.000	Class 2 vs. 4	257.457	0.000
Class 3 vs. 4	55.335	0.000			

In this portion of the OUTPUT, Mplus reports the Means and Standard Errors (SE) of variable **happy** across the 4 latent classes estimated in the model. Participants in the “Optimist” class (Class 2) report the highest average in happiness, 8.10 (SE = 0.04), while those in the “Sceptical” class (Class 4) report the lowest average: 7.24 (SE = 0.04).

The **Chi-Square** tests and relative *p* values (*df*=1) indicate that the differences in happiness between “Optimists” and the other latent classes are significant:

Class 1 vs. 2 =98.72;

Class 2 vs. 3 = 73.31;

Class 2 vs. 4 = 257.46.