

Solutions to 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 **rspppsgv ractrolg rspppi1 rcpttpol rptcppl retapap1**, 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**.

ANSWER:

See file “ess_ex4_1.inp”. The key options are here:

Data:

File is ess_ex3.dat ;

Variable:

Names are

essround idno polintr happy rspppsgv ractrolg rspppi1 rcpttpol rptcppl
retapap1 cou ess7id nuts1en nuts2en nuts3en agec edu male dcou1 dcou2
dcou3 dcou4 dcou5 dedu1 dedu2 dedu3;

Missing are all (-999) ;

Usevariables are

rspppsgv ractrolg rspppi1 rcpttpol rptcppl retapap1;

CATEGORICAL= rspppsgv ractrolg rspppi1 rcpttpol rptcppl retapap1;

Classes=class(4);

IDVAR=ess7id;

AUXILIARY= happy agec edu male dcou1 dcou2
dcou3 dcou4 dcou5 dedu1 dedu2 dedu3;

SAVEDATA:

File=ess4cl.dat;

SAVE=CPROB;

MISSFLAG=-999;

In particular, this line in command **VARIABLE:**

AUXILIARY= happy agec edu male dcou1 dcou2 dcou3 dcou4 dcou5 dedu1 dedu2 dedu3;

instructs Mplus to include these “auxiliary” variables in the datafile that will be created, but at the same time the variables are not taken into consideration in the model estimation.

These lines:

SAVEDATA:

File=ess4cl.dat;

SAVE=CPROB;

MISSFLAG=-999;

instruct Mplus to save a datafile at the end of the model estimation; the datafile is going to be called “ess4cl.dat” (you can save the file in other formats, e.g. .txt).

SAVE=CPROB instructs Mplus to save the posterior probabilities of being in the estimated latent classes, and, as default, Mplus will also save the *most likely latent class membership*.

Finally, the option **MISSFLAG=-999**; instructs Mplus to assign value -999 to any value missing (you can select different values or characters).

Note that by including line

```
IDVAR=ess7id;
```

in command **VARIABLE**: I have instructed Mplus to save the IDs of participants in the datafile, which is useful when you want to check the file or match-merge it with other files.

The information about the datafile saved by Mplus is provided at the end of the output, see “ess_ex4_1.out”:

SAVEDATA INFORMATION

Save file
ess4cl.dat

Order and format of variables

RPSPPSGV	F10.3
RACTROLG	F10.3
RPSPPIPL	F10.3
RCPTPOL	F10.3
RPTCPPLT	F10.3
RETAPAPL	F10.3
ESS7ID	F6.0
HAPPY	F10.3
AGEC	F10.3
EDU	F10.3
MALE	F10.3
DCOU1	F10.3
DCOU2	F10.3
DCOU3	F10.3
DCOU4	F10.3
DCOU5	F10.3
DEDU1	F10.3
DEDU2	F10.3
DEDU3	F10.3
CPROB1	F10.3
CPROB2	F10.3
CPROB3	F10.3
CPROB4	F10.3
CLASS	F10.3

This information is essential to use the datafile produced. Note that the last variables in the file are the posterior probabilities of being in latent class 1, 2, 3 and 4 respectively, and the most likely class: the latter is called “**class**” because that is what I called the latent class, but you can call the latent class any name you want (e.g. “c”, “engage”, etc.).

This output and the datafile created correspond to STEP 1 of the Three-Step Approach.

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.

ANSWER:

Firstly, we need to extract the information about measurement error in the model with 4 classes we have just estimated. This information is in file “ess_ex4_1.out”:

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent Class (Column)				
	1	2	3	4
1	0.848	0.022	0.071	0.059
2	0.033	0.894	0.073	0.000
3	0.054	0.058	0.842	0.046
4	0.033	0.000	0.053	0.914

Classification Probabilities for the Most Likely Latent Class Membership (Row) by Latent Class (Column)				
	1	2	3	4
1	0.843	0.027	0.069	0.061
2	0.027	0.882	0.091	0.000
3	0.054	0.045	0.827	0.074
4	0.032	0.000	0.033	0.934

Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Row) by Latent Class (Column)				
	1	2	3	4
1	2.630	-0.803	0.134	0.000
2	4.430	7.915	5.645	0.000
3	-0.316	-0.500	2.412	0.000
4	-3.363	-9.530	-3.340	0.000

The last table with the Logits for the Classification Probabilities is the one we will be using to fix the measurement parameters of the model. These logits provide information about uncertainty in latent class membership in the model we have just estimated.

Extracting this information corresponds to STEP 2 of the Three-Step Approach.

With this information, we can now write an INPUT file in which we instruct Mplus to use the most likely class membership from the previous latent class estimation as a nominal indicator of latent class, which is then used to estimate the 4 latent classes. The association between these four latent classes and the indicators is going to be fixed to the measurement error parameters, the logits for the classification probabilities we have inspected above.

The INPUT file is “ess_ex4_2.inp”.

Let's consider the commands FILE: and VARIABLE: first:

```

Data:
  File = ess4cl.dat ;
Variable:
  Names =
  rpspsgv ractrolg rpsppi1 rcptppol rptcppl1 retapap1
  ess7id
  happy agec edu male dcou1 dcou2
    dcou3 dcou4 dcou5 dedu1 dedu2 dedu3
    cprob1 cprob2 cprob3 cprob4 class;
Missing are all (-999) ;
USEVAR=agec class male dcou1 dcou2
    dcou3 dcou4 dedu2 dedu3 ;
Classes=newclass(4);

NOMINAL=class;

```

The file "ess4cl.dat" is the one we asked Mplus to produce in the previous model estimation (Step 1).

In the INPUT, we need to instruct Mplus to name the variables in the order they appear: this order was specified in the OUTPUT reported in Page 2.

It is pivotal to instruct Mplus to read missing value indicators (-999) as missing values, lest Mplus will consider these as valid values.

The following line:

```
USEVAR=agec class male dcou1 dcou2 dcou3 dcou4 dedu2 dedu3 ;
```

instructs Mplus to consider these as variables in the model. The command **MODEL:** will specify how these variables are going to be used in the model.

The following line:

```
Classes=newclass(4);
```

instructs Mplus to estimate a latent class variable (which I called "**newclass**") with 4 classes.

The line:

```
NOMINAL=class;
```

instructs Mplus to consider the "**class**" variable, which represents participants' most likely class membership, as a nominal variable (i.e. non-ordered categories). In the 3-step approach, the most-likely class is used as an indicator of the latent class, as we will specify in the **MODEL:** command.

The MODEL: command is copied here:

```
MODEL:
```

```
%OVERALL%
```

```
newclass ON male agec dcou1 dcou2
```

```
dcou3 dcou4 dedu2 dedu3;
```

```
%newclass#1%
[class#1@2.630];
[class#2@-0.803];
[class#3@0.134];
```

```
%newclass#2%
[class#1@4.430];
[class#2@7.915];
[class#3@5.645];
```

```
%newclass#3%
[class#1@-0.316];
[class#2@-0.500];
[class#3@2.412];
```

```
%newclass#4%
[class#1@-3.363];
[class#2@-9.530];
[class#3@-3.340];
```

The first lines are

```
%OVERALL%
newclass ON male agec dcou1 dcou2 dcou3 dcou4 dedu2 dedu3;
```

What follows **%OVERALL%** is the general part of the model, the one that will apply to *all* 4 classes of the latent class model being estimated.

In the statement above, I am asking Mplus to estimate a (multinomial logistic) regression of latent classes on the covariates (dummy variables that represent *gender, age, country, educational attainment*). As in previous examples, note the use of dummy variables, where the omission of one dummy variable allows this as the reference category. For example, by excluding the dummy variable associated with “France” (**dcou5**), I am considering France as the reference category in the analyses.

The lines:

```
%newclass#1%
...
%newclass#2%
...
```

specify the *class-specific model parameters*.

For example:

```
%newclass#1%
```

```
[class#1@2.630];
[class#2@-0.803];
[class#3@0.134];
```

specify the parameters specific to latent class 1 (**#1**) of the latent class “**newclass**”.

The statements that follow **%newclass#1%** are instructing Mplus to consider the nominal variable **class** (see **VARIABLE:** command) as the latent class indicator. Since this indicator has 4 categories, we need to specify the log odds of being in category 1 of **class** rather than being in category 4 (the reference category: remember that Mplus by default considers the last category in multinomial regression analyses as the reference category), and so on.

These are fixed to the measurement error estimated in Step 2, i.e. to the logit values reported by the output of the estimated measurement model, those reported in Page 3. As you can see, the values of **[class#1]**, **[class#2]** and **[class#3]** in **%newclass#1%** are all fixed to the values reported in the first row of the Table “Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Row) by Latent Class (Column)” reported in Page 3. Similarly, I have used the values from the other rows in the Table in Page 3 to fix the association between indicators and latent class in **%newclass#2%**, and so on.

3. Check the solution of Task 2 above. Do the latent classes estimated correspond in number to those estimated in Task 1?

Compare the output “ess_ex4_1.out” and “ess_ex4_2.out”.

There is a difference in the number of participants in each analysis. While the latent class model had been estimated on $N = 8921$, the model with latent classes and covariates includes $N = 8883$. You can investigate why by including option **SAMPSTAT** (i.e. report sample summary statistics) in the command **OUTPUT:** In particular, checking the pattern of missing data will indicate that the $N = 38$ excluded from the model in “ess_ex4_2.out” display missing data in one or more of the covariates, and are therefore excluded in the model where latent class is regressed on the covariates.

Note: in this example, there is a low rate of missing values on the covariates. However, if the problem was more substantial, you might want to tackle this issue using maximum likelihood estimation or multiple imputation.

You can check that the number of individuals assigned to the classes of **newclass** correspond to the numbers assigned to classes of the originally estimated **class** variable, save for the exclusion of those with missing data in the covariates:

Output from “ess_ex4_1.out”

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

Latent Classes		
1	1795	0.20121
2	1470	0.16478
3	2333	0.26152
4	3323	0.37249

Output from "ess_ex4_2.out":

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

: Class Counts and Proportions

Latent Classes		
1	1791	0.20162
2	1465	0.16492
3	2324	0.26162
4	3303	0.37183

And by obtaining a Datafile with the most likely class membership for **newclass** in Step 3, we can compare latent class membership of participants with latent class allocation obtained in Step 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.

Check the Output file "ess_ex4_2.out".

In "ess_ex4_2.out" solution, the "Sceptical" class is latent class 4, which, being the last one, is also the reference class in Mplus. Latent class 2 is the "Optimist" latent class (refer to output from Step 1, "ess_ex4_1.out" to check the conditional item probabilities by class).

Note that in the **OUTPUT:** command of the input file I had included option **CINTERVALS**, which instructs Mplus to report confidence intervals of the parameters.

Mplus reports Odds Ratios of logistic regressions:

LOGISTIC REGRESSION ODDS RATIO RESULTS

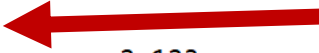
Categorical Latent Variables

NEWCLASS ON

MALE	2.037
AGEC	0.984
DCOU1	1.197
DCOU2	0.589
DCOU3	1.155
DCOU4	1.327
DEDU2	2.166
DEDU3	6.142

NEWCLASS ON

MALE	2.123
AGEC	0.983
DCOU1	1.849
DCOU2	1.327
DCOU3	16.646
DCOU4	0.870
DEDU2	1.222
DEDU3	6.838



Odds Ratios for
NEWCLASS # 2

We can check the Odds Ratios for the Country dummy variables in **newclass** #2: these are the odds ratios of being in latent class 2 (“Optimistic”) rather than latent class 4 (“Sceptical”). The missing Country dummy variable (**dcou5**, i.e. France) is the reference category. Therefore, the odds ratios indicate the change in the odds of being in the “Optimist” rather than the “Sceptical” class for each country, compared to “France”.

Since country 1 = Austria, country 2 = Belgium, country 3 = Switzerland, country 4 = Spain, we can report the odds ratios as follows:

Country	Odds ratios of Optimist rather than Sceptic
Austria	1.85
Belgium	1.33
Switzerland	16.65
Spain	0.87
France	<i>reference</i>

The confidence intervals can be found in another table (provided you have included the option **CINTERVALS** in the **OUTPUT:** command):

CONFIDENCE INTERVALS FOR THE LOGISTIC REGRESSION ODDS RATIO RESULTS

Categorical Latent Variables

NEWCLASS ON							
MALE	1.672	1.752	1.795	2.037	2.311	2.367	2.482
AGEC	0.979	0.980	0.980	0.984	0.987	0.987	0.989
DCOU1	0.904	0.967	1.001	1.197	1.432	1.482	1.585
DCOU2	0.427	0.461	0.480	0.589	0.72	0.753	0.813
DCOU3	0.731	0.816	0.862	1.155	1.448	1.637	1.826
DCOU4	1.005	1.074	1.111	1.327	1.586	1.641	1.753
DEDU2	1.669	1.776	1.834	2.166	2.559	2.642	2.812
DEDU3	4.603	4.931	5.108	6.142	7.384	7.649	8.195
NEWCLASS ON							
MALE	1.720	1.809	1.856	2.123	2.428	2.491	2.619
AGEC	0.977	0.979	0.979	0.983	0.987	0.987	0.989
DCOU1	1.292	1.407	1.471	1.849	2.326	2.430	2.648
DCOU2	0.934	1.015	1.060	1.327	1.662	1.735	1.887
DCOU3	11.398	12.478	13.070	16.646	21.201	22.206	24.310
DCOU4	0.589	0.647	0.678	0.870	1.115	1.170	1.284

Values and their CI
for newclass # 2

Mplus has the annoying habit of not reporting the headings across tables, so to make sense of this table you should look at the headings of the previous one:

CONFIDENCE INTERVALS OF MODEL RESULTS

Lower .5% Lower 2.5% Lower 5% Estimate Upper 5% Upper 2.5% Upper .5%

You can now see that the first value in the table is the lower 0.5% Confidence Interval (CI), then the lower 2.5% CI, and so on.

In this way, we could take these values from the table:

NEWCLASS ON							
MALE	1.720	1.809	1.856	2.123	2.428	2.491	2.619
AGEC	0.977	0.979	0.979	0.983	0.987	0.987	0.989
DCOU1	1.292	<u>1.407</u>	1.471	1.849	2.326	<u>2.430</u>	2.648
DCOU2	0.934	<u>1.015</u>	1.060	1.327	1.662	<u>1.735</u>	1.887
DCOU3	11.398	<u>12.478</u>	13.070	16.646	21.201	<u>22.206</u>	24.310
DCOU4	0.589	<u>0.647</u>	0.678	0.870	1.115	<u>1.170</u>	1.284
DEDU2	0.922	0.986	1.020	1.222	1.463	1.514	1.620
DEDU3	5.099	5.469	5.669	6.838	8.247	8.548	9.170

and report them in a table like this:

Country	Odds ratios of Optimist rather than Sceptic	95% Confidence Intervals
Austria	1.85	1.41 to 2.43
Belgium	1.33	1.02 to 1.74
Switzerland	16.65	12.45 to 22.21
Spain	0.87	0.65 to 1.17
France	reference	---

The results suggest that, compared to participants in France, participants in Switzerland displayed almost a 17-fold increase in the odds of being “Optimist” rather than “Sceptical” regarding people’s ability to influence politics, an encouraging endorsement of the Swiss way to democracy!

If you wanted to check the p values of these results, these are reported together with the parameter estimates:

Categorical Latent Variables

NEWCLASS#1 ON

MALE	0.711	0.077	9.273	0.000
AGEC	-0.017	0.002	-8.473	0.000
DCOU1	0.180	0.109	1.650	0.099
DCOU2	-0.529	0.125	-4.226	0.000
DCOU3	0.144	0.178	0.812	0.417
DCOU4	0.283	0.108	2.620	0.009
DEDU2	0.773	0.101	7.633	0.000
DEDU3	1.815	0.112	16.212	0.000

NEWCLASS#2 ON

MALE	0.753	0.082	9.220	0.000
AGEC	-0.017	0.002	-7.502	0.000
DCOU1	0.615	0.139	4.412	0.000
DCOU2	0.283	0.137	2.072	0.038
DCOU3	2.812	0.147	19.127	0.000
DCOU4	-0.139	0.151	-0.923	0.356
DEDU2	0.200	0.109	1.829	0.067
DEDU3	1.922	0.114	16.874	0.000

Once again, to find out what these numbers represent, we have to look for the headings in previous tables:

MODEL RESULTS

Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
----------	------	-----------	-----------------------

In this way, we can check that the differences in the odds of being optimist rather than sceptical are not significantly different when we compare France and Spain ($p = 0.356$), but are significant when we compare France and the remaining countries.

Note that we could have extracted the Odds Ratios of these comparisons from the table with the “Estimate” parameters: since these are reported in log odds, we can obtain the odds by exponentiating. For example, by exponentiating the estimate of the comparison between Switzerland (dcou3) and France (the reference category), we would obtain $\exp(2.812) = 16.643$, which, allowing for differences in rounding, is the Odds Ratio we have already reported.

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 adjusted mean and variance of happy across the 4 classes;
 - c. Test whether average **happy** is similar in latent class “Optimist” and “Sceptical”.

Check INPUT file “ess_ex4_3.inp” and OUTPUT file “ess_ex4_3.out”.

Solution to (a):

The **MODEL:** command includes these statements:

```
MODEL:
%OVERALL%
newclass ON male agec dcou1 dcou2 dcou3 dcou4 dedu2 dedu3;
happy ON male agec dcou1 dcou2 dcou3 dcou4 dedu2 dedu3;
```

The last line above instructs Mplus to model the regression of distal outcome **happy** on the other covariates.

Solution to (b):

The input file also includes these options in the **MODEL:** command (highlighted in Bold):

```
%newclass#1%
[class#1@2.630];
[class#2@-0.803];
[class#3@0.134];
[happy] (p1);
happy*;

%newclass#2%
[class#1@4.430];
[class#2@7.915];
[class#3@5.645];
[happy](p2);
happy*;

%newclass#3%
[class#1@-0.316];
[class#2@-0.500];
[class#3@2.412];
[happy] (p3);
happy*;

%newclass#4%
```

```
[class#1@-3.363];
[class#2@-9.530];
[class#3@-3.340];
[happy] (p4);
happy*;
```

These options are instructing Mplus to estimate the mean of distal outcome variable **happy** within each latent class of **newclass**. Remember that the square parentheses **[]** indicate the mean of a continuous variable in Mplus. Thus, mentioning **[happy]** within **%newclass#1%**, **%newclass#2%**, etc. is instructing Mplus to estimate the mean of **happy** within **newclass=1**, **newclass=2**, etc.

Note that each **[happy]** is followed by a sequence of letters and number between round parentheses, **(p1)**, **(p2)** etc. This allows to name these parameters and use these names to instruct Mplus to constrain them or to test hypotheses about them, as I will explain in Solution to (c).

In order to estimate the variance of **happy** within each class, I also included the option **happy*** within each class **%newclass#1%**, etc. The name of continuous variable without parentheses is how Mplus refers to variances, and the asterisk ***** indicates that parameter is freed. Consider that the overall model is also estimating the regression of **happy** on other covariates, so the variance being estimated here is the adjusted variance.

In the OUTPUT "ess_ex4_3.out" you will find the estimates of (adjusted) means of happiness within the four classes:

	Estimate	S.E.	Est./S.E.	P-Value
Latent Class 1				
HAPPY ON				
MALE	-0.126	0.034	-3.684	0.000
AGEC	-0.004	0.001	-4.879	0.000
DCOU1	0.196	0.059	3.340	0.001
DCOU2	0.417	0.050	8.351	0.000
DCOU3	0.518	0.055	9.435	0.000
DCOU4	0.348	0.057	6.116	0.000
DEDU2	0.129	0.045	2.867	0.004
DEDU3	0.202	0.048	4.245	0.000
Means				
CLASS#1	2.630	0.000	999.000	999.000
CLASS#2	-0.803	0.000	999.000	999.000
CLASS#3	0.134	0.000	999.000	999.000
Intercepts				
HAPPY	7.019	0.100	69.919	0.000
Residual Variances				
HAPPY	3.830	0.276	13.873	0.000

The results from the output are:

Latent Class	Adjusted Mean	Adjusted Variance
1 (Disenchanted)	7.02	3.83
2 (Optimist)	7.92	1.37
3 (Neutral)	7.60	1.43
4 (Sceptic)	6.93	3.97

Solution to (c):

In INPUT “ess_ex4_3.inp” I have asked Mplus to estimate the (adjusted) means of happiness in the four latent classes (see solution to (b) above). I have also given names to these parameters, so that the mean of happy in latent class 2 (“Optimist”) is called **p2**, and the one in latent class 4 (“Sceptic”) is called **p4**.

After the MODEL: command, I also included a **MODEL TEST:** command:

MODEL TEST:

p2=p4;

This is asking Mplus to test the null hypothesis that parameter **p2** (the adjusted average happiness of individuals in latent class 2) is equal to parameter **p4** (the adjusted average happiness of individuals in latent class 4). Mplus will provide a Wald test of equality constraints.

In the OUTPUT “ess_ex4_3.out” you will find this statement:

Wald Test of Parameter Constraints

Value	201.061
Degrees of Freedom	1
P-Value	0.0000

This indicates that the null hypotheses of the average happiness being equal for those in latent class 2 and those in latent class 4 should be rejected.