Appendix C.III: Hypothesized Method - Confirmatory Latent Class Analysis (CLCA)

Adam Garber

2024-08-20

```
library(tidyverse)
library(MplusAutomation)
library(rhdf5)
library(here)
library(glue)
library(gt)
library(janitor)
library(reshape2)
library(cowplot)
```

Read data file `n_6000_lca_rep1.dat` (N=6000; Replication 5)

```
C3_data <- read.delim2(here("C1-Simulation", "n_6000_lca_rep5.dat"), sep = "", header = FALSE) %>%
    select(-V18) %>%
    setNames(c("primary", "change", "interupt", "initiat", "engage", "approach",
        "response", "expect", "new", "same", "relative", "objects", "sequence", "trans",
        "avoid", "control", "touch")) %>%
    purrr::modify_if(is.character, as.numeric)
```

```
write_csv(C3_data, here("data", "C3_simulated_N6000R5.csv"))
```

Procedure: Hypothesized Model

- Step 1: Estimate unconstrained model (comparison model)
- Step 2: Estimate hypothesis-driven model with boundary and inequality constraints

---

Step 1

Estimate unconstrainded LCA Model with Simulated Data (`C3_data`)

```
m_unconstrained  <- mplusObject(

  TITLE = "Unconstrained Model",

  VARIABLE =
   "categorical = primary-touch;
    usevar = primary-touch;
    classes = c(5); ",

  ANALYSIS =
   "estimator = mlr;
    type = mixture;
    starts = 0;
    !starts = 500 200;
    STSEED = 21345; !!! USE SEED TO REPLICATE THESIS RESULTS !!!
   ",

  OUTPUT = "svalues;",

  PLOT = "type = plot3;
          series = primary-touch(*);",

  usevariables = colnames(C3_data),
  rdata = C3_data)

m_unconstrained_fit <- mplusModeler(m_unconstrained,
                 dataout=here("C3-Hypothesized", "Unconstrained.dat"),
                 modelout=here("C3-Hypothesized", "Unconstrained.inp") ,
                 check=TRUE, run = TRUE, hashfilename = FALSE)
```

Step 2

Estimate Revised Hypothesis Model

```
m1  <- mplusObject(

  TITLE = "Hypothesis-M1",

  VARIABLE =
   "categorical = primary-touch;
    usevar = primary-touch;
    classes = c(5); ",

  ANALYSIS =
   "estimator = mlr;
    type = mixture;
    !starts = 500 200;
    !STSEED = 459573; !!! USE SEED TO REPLICATE THESIS RESULTS !!!
```

```
  ",

MODEL =
  "!!! NAME & LABEL PARAMETERS !!!

   %OVERALL%

   %C#1%       !!!High_Flex!!!

   [ primary$1*-.85 ](t1);
   [  change$1*-.85 ](t2);
   [interupt$1*-.85 ](t3);
   [ initiat$1*-.85 ](t4);
   [  engage$1*-.85 ](t5);
   [approach$1*-.85 ](t6);
   [response$1*-.85 ](t7);
   [  expect$1*-.85 ](t8);
   [     new$1*-.85 ](t9);
   [    same$1*-.85 ](t10);
   [relative$1*-.85 ](t11);
   [ objects$1*-.85 ](t12);
   [sequence$1*-.85 ](t13);
   [   trans$1*-.85 ](t14);
   [   avoid$1*-.85 ](t15);
   [ control$1*-.85 ](t16);
   [   touch$1*-.85 ](t17);

   %C#2%       !!!ScIn_Flex!!!

   [ primary$1*-.85 ](t18);
   [  change$1*-.85 ](t19);
   [interupt$1*-.85 ](t20);
   [ initiat$1*-.85 ](t21);
   [  engage$1*-.85 ](t22);
   [approach$1*-.85 ](t23);
   [response$1*-.85 ](t24);
   [  expect$1*-.85 ](t25);
   [     new$1*.85  ](t26);
   [    same$1*.85  ](t27);
   [relative$1*.85  ](t28);
   [ objects$1*.85  ](t29);
   [sequence$1*.85  ](t30);
   [   trans$1*.85  ](t31);
   [   avoid$1*.85  ](t32);
   [ control$1*.85  ](t33);
   [   touch$1*.85  ](t34);

   %C#3%        !!!LocSnsLo!!!

  [ primary$1*.85 ](t35);
  [  change$1*.85 ](t36);
  [interupt$1*.85 ](t37);
  [ initiat$1*.85 ](t38);
```

```
[  engage$1*.85 ](t39);
[approach$1*.85 ](t40);
[response$1*.85 ](t41);
[  expect$1*.85 ](t42);
[     new$1*-.85](t43);
[    same$1*-.85](t44);
[relative$1*-.85](t45);
[ objects$1*-.85](t46);
[sequence$1*-.85](t47);
[   trans$1*-.85](t48);
[   avoid$1*-.85](t49);
[ control$1*-.85](t50);
[   touch$1*-.85](t51);

 %C#4%     !!!Env_Flex !!!

    [ primary$1 ](t52);
    [  change$1 ](t53);
    [interupt$1 ](t54);
    [ initiat$1 ](t55);
    [  engage$1 ](t56);
    [approach$1 ](t57);
    [response$1 ](t58);
    [  expect$1 ](t59);
[     new$1*.85 ](t60);
[    same$1*.85 ](t61);
[relative$1*.85 ](t62);
[ objects$1*-.85](t63);
[sequence$1*-.85](t64);
[   trans$1*-.85](t65);
[   avoid$1*.85 ](t66);
[ control$1*.85 ](t67);
[   touch$1*.85 ](t68);

 %C#5%         !!!Low_Flex!!!

 [ primary$1*.85 ](t69);
 [  change$1*.85 ](t70);
 [interupt$1*.85 ](t71);
 [ initiat$1*.85 ](t72);
 [  engage$1*.85 ](t73);
 [approach$1*.85 ](t74);
 [response$1*.85 ](t75);
 [  expect$1*.85 ](t76);
 [     new$1*.85 ](t77);
 [    same$1*.85 ](t78);
 [relative$1*.85 ](t79);
 [ objects$1*.85 ](t80);
 [sequence$1*.85 ](t81);
 [   trans$1*.85 ](t82);
 [   avoid$1*.85 ](t83);
 [ control$1*.85 ](t84);
 [   touch$1*.85 ](t85);
```

```
  ",

MODELCONSTRAINT = "
!!!       THRESHOLD BOUNDARIES            !!!
!!! -.85 (THRESHOLD) ~ .70 (PROBABILTY)   !!!
!!!  .85 (THRESHOLD) ~ .30 (PROBABILTY)   !!!

! LABELS C1-C5 BELOW REFLECT ORDER OF CLASSES IN PLOT (NOT MPLUS C# LABELS ABOVE)

   !   C1    |   C2    |   C3    |   C4   |    C5  !
   !High_Flex|ScIn_Flex|Env_Flex |LocSnsLo|Low_Flex!
   !_________|_________|_________|________|________!
    t1 <-.8;  t18<-.8;  t35> .8;           t69> .8;  !!!U1 !!!
    t2 <-.8;  t19<-.8;  t36> .8;           t70> .8;  !!!U2 !!!
    t3 <-.8;  t20<-.8;  t37> .8;           t71> .8;  !!!U3 !!!
    t4 <-.8;  t21<-.8;  t38> .8;           t72> .8;  !!!U4 !!!
    t5 <-.8;  t22<-.8;  t39> .8;           t73> .8;  !!!U5 !!!
    t6 <-.8;  t23<-.8;  t40> .8;           t74> .8;  !!!U6 !!!
    t7 <-.8;  t24<-.8;  t41> .8;           t75> .8;  !!!U7 !!!
    t8 <-.8;  t25<-.8;  t42> .8;           t76> .8;  !!!U8 !!!

    t9 <-.8;  t26> .8;  t43<-.8; t60> .8;  t77> .8;  !!!U9 !!!
    t10<-.8;  t27> .8;  t44<-.8; t61> .8;  t78> .8;  !!!U10!!!
    t11<-.8;  t28> .8;  t45<-.8; t62> .8;  t79> .8;  !!!U11!!!
    t12<-.8;  t29> .8;  t46<-.8; t63<-.8;  t80> .8;  !!!U12!!!
    t13<-.8;  t30> .8;  t47<-.8; t64<-.8;  t81> .8;  !!!U13!!!
    t14<-.8;  t31> .8;  t48<-.8; t65<-.8;  t82> .8;  !!!U14!!!
    t15<-.8;  t32> .8;  t49<-.8; t66> .8;  t83> .8;  !!!U15!!!
    t16<-.8;  t33> .8;  t50<-.8; t67> .8;  t84> .8;  !!!U16!!!
    t17<-.8;  t34> .8;  t51<-.8; t68> .8;  t85> .8;  !!!U17!!!

!!!       EQUALITY (& INEQUALITY) CONSTRAINTS       !!!

  !  C1vC2  |  C3vC4  |  C4vC5  !
  !_________|_________|_________!
             t35=-t52;           !!!U1 !!!
             t36=-t53;           !!!U2 !!!
             t37=-t54;           !!!U3 !!!
             t38=-t55;           !!!U4 !!!
             t39=-t56;           !!!U5 !!!
             t40=-t57;           !!!U6 !!!
             t41=-t58;           !!!U7 !!!
             t42=-t59;           !!!U8 !!!
   t9 =-t26;                     !!!U9 !!!
   t10=-t27;                     !!!U10!!!
   t11=-t28;                     !!!U11!!!
   t12=-t29;           t63=-t80; !!!U12!!!
   t13=-t30;           t64=-t81; !!!U13!!!
   t14=-t31;           t65=-t82; !!!U14!!!
   t15=-t32;                     !!!U15!!!
   t16=-t33;                     !!!U16!!!
   t17=-t34;                     !!!U17!!!
```

```
  ",

  OUTPUT = "tech14;",

  PLOT = "type = plot3;
          series = primary-touch(*);",

  usevariables = colnames(C3_data),
  rdata = C3_data)

m1_fit <- mplusModeler(m1,
                  dataout=here("C3-Hypothesized", "Hypothesis-CLCA.dat"),
                  modelout=here("C3-Hypothesized", "Hypothesis-CLCA.inp") ,
                  check=TRUE, run = TRUE, hashfilename = FALSE)
```

Step 3: Compare model fit

Conduct the Sattorra-Bentler adjusted Log Likelihood Ratio (LRT) difference test:

- Baseline model (parent): `Unconstrained.out` is the 'un-constrained model" with 89 parameters.
- Confirmatory model (nested): `Hypothesis.CLCA.out` is the "constrained model" with 69 parameters.

```
hypothesized_models <- readModels(here("C3-Hypothesized"), quiet = TRUE)
```

```
# *0 = null or nested model & *1 = comparison or parent model

# Log Likelihood Values
L0 <- hypothesized_models[["Hypothesis.CLCA.out"]][["summaries"]][["LL"]]
L1 <- hypothesized_models[["Unconstrained.out"]][["summaries"]][["LL"]]

# LRT equation
lr <- -2 * (L0 - L1)

# Parameters
p0 <- hypothesized_models[["Hypothesis.CLCA.out"]][["summaries"]][["Parameters"]]
p1 <- hypothesized_models[["Unconstrained.out"]][["summaries"]][["Parameters"]]

# Scaling Correction Factors
c0 <- hypothesized_models[["Hypothesis.CLCA.out"]][["summaries"]][["LLCorrectionFactor"]]
c1 <- hypothesized_models[["Unconstrained.out"]][["summaries"]][["LLCorrectionFactor"]]

# Difference Test Scaling correction (Sattorra-Bentler adjustment)
cd <- ((p0 * c0) - (p1 * c1))/(p0 - p1)

# Chi-square difference test(TRd)
TRd <- (lr)/(cd)
```

```r
# Degrees of freedom
df <- abs(p0 - p1)


# Significance test
(p_diff <- pchisq(TRd, df, lower.tail = FALSE))
```

```
## [1] 0
```

RESULT: The Log Likelihood χ^2 difference test comparing the Baseline and Hypothesized LCA models was, $\chi^2(20) = 2929.14, p < .001$. See Reference Here

---

Compare model fit summary statistics: Unconstrained and Hypothesized

```r
enum_extract1 <- LatexSummaryTable(hypothesized_models$Unconstrained.out, keepCols = c("Title",
    "Parameters", "LL", "BIC", "aBIC", "Observations"))
enum_extract2 <- LatexSummaryTable(hypothesized_models$Hypothesis.CLCA.out, keepCols = c("Title",
    "Parameters", "LL", "BIC", "aBIC", "Observations"))
```

Calculate indices derived from the Log Likelihood (LL)

```r
allFit <- rbind(enum_extract1, enum_extract2) %>%
    mutate(aBIC = -2 * LL + Parameters * log((Observations + 2)/24)) %>%
    select(1:5) %>%
    mutate(Title = case_when(Title == " Unconstrained Model" ~ "Unconstrained Model",
        Title == " Hypothesis-M1" ~ "Hypothesized Model"))
```

Format fit table

```r
allFit %>%
    gt() %>%
    tab_header(title = md("**Model Fit Comparision Table**"), subtitle = md(" ")) %>%
    cols_label(Title = "Model", Parameters = md("Par"), LL = md("*LL*"), BIC = md("BIC"),
        aBIC = md("aBIC")) %>%
    tab_footnote(footnote = md("*Note.* Par = Parameters; *LL* = model log likelihood;
     BIC = Bayesian information criterion; aBIC = sample size adjusted BIC."),
        locations = cells_title()) %>%
    tab_options(column_labels.font.weight = "bold")
```

Model Fit Comparision Table[1]

[1]

Model	Par	*LL*	BIC	aBIC
Unconstrained Model	89	-52601.37	105977.0	105694.2
Hypothesized Model	69	-54038.40	108677.1	108457.8

[1] *Note.* Par = Parameters; LL = model log likelihood; BIC = Bayesian information criterion; aBIC = sample size adjusted BIC.