Creating Summary Fit Tables for LCA and LTA Analyses Using `MplusAutomation`

Adam Garber

September, 02, 2021

This syntax automates the process of generating a model fit summary table for comparing a series of LCA models generated during enumeration. The code produces a publication read table that approximately adheres to APA formatting guidelines. Included in the table is the exhaustive set of fit indices recommended as current best practice for making enumeration decisions based on simulation results (Nylund et al., 2007). A separate example is provided below to demonstrate the procedure for an LTA analysis with two time points.

If using this tutorial to produce tables for publication it would be greatly appreciated if you cite this resource using the citation provided here:

Garber, A. C. (2021). Creating Summary Fit Tables for LCA and LTA Analyses Using MplusAutomation. Retrieved from psyarxiv.com/uq2fh

---

Associated Github repository here:

https://github.com/garberadamc/Fit-Summary-Table-LCA-LTA

Load packages

```
library(MplusAutomation) # Conduit between R & Mplus
library(tidyverse)       # Data manipulation
library(here)            # Location, location, location
library(gt)              # Tables with seamless knitting
```

Create Model Fit Summary Table for Latent Class Analysis (LCA)

STEP 1. Read in models from location where enumeration files are stored (e.g., `enum_mplus` sub-folder)

```
output_enum <- readModels(here("enum_mplus"), quiet = TRUE)
```

STEP 2. Extract relevant fit statistics from `mplusObject()` object (e.g., named `output_enum` in this example)

```
enum_extract <- LatexSummaryTable(output_enum,
                keepCols=c("Title", "Parameters", "LL", "BIC", "aBIC",
                           "BLRT_PValue", "T11_VLMR_PValue","Observations"))
```

STEP 3. Calculate additional fit indices derived from the Log Likelihood (LL) as recommended in Nylund et al., 2007.

```
allFit <- enum_extract %>%
  mutate(aBIC = -2*LL+Parameters*log((Observations+2)/24)) %>%
  mutate(CIAC = -2*LL+Parameters*(log(Observations)+1)) %>%
  mutate(AWE = -2*LL+2*Parameters*(log(Observations)+1.5)) %>%
  mutate(SIC = -.5*BIC) %>%
  mutate(expSIC = exp(SIC - max(SIC))) %>%
  mutate(BF = exp(SIC-lead(SIC))) %>%
  mutate(cmPk = expSIC/sum(expSIC)) %>%
  select(1:5,9:10,6:7,13,14) %>%
  arrange(Parameters)
```

STEP 4. Generate and format the model fit table using the {`gt`} package.

```
allFit %>%
  mutate(Title = str_remove(Title, " LCA Enumeration - Youth Coping Strategies")) %>%
  gt() %>%
  tab_header(
    title = md("**Model Fit Summary Table**"), subtitle = md(" ")) %>%
  cols_label(
    Title = "Classes",
    Parameters = md("Par"),
    LL = md("*LL*"),
    T11_VLMR_PValue = "VLMR",
    BLRT_PValue = "BLRT",
    BF = md("BF"),
    cmPk = md("*cmP_k*")) %>%
  tab_footnote(
    footnote = md(
    "*Note.* Par = parameters; *LL* = log likelihood;
      BIC = bayesian information criterion;
      aBIC = sample size adjusted BIC;
```

```
    CAIC = consistent Akaike information criterion;
    AWE = approximate weight of evidence criterion;
    BLRT = bootstrapped likelihood ratio test p-value;
    VLMR = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test p-value;
    cmPk = approximate correct model probability."),
  locations = cells_title()) %>%
tab_options(column_labels.font.weight = "bold") %>%
fmt_number(10,decimals = 2,
           drop_trailing_zeros=TRUE,
           suffixing = TRUE) %>%
fmt_number(c(3:9,11),
           decimals = 1) %>%
fmt_missing(1:11,
            missing_text = "--") %>%
fmt(c(8:9,11),
  fns = function(x)
  ifelse(x < 0.001, "<.001",
         scales::number(x, accuracy = 0.01))) %>%
fmt(10, fns = function(x)
  ifelse(x>100, ">100",
         scales::number(x, accuracy = .1)))
```

Model Fit Summary Table[1]

Classes	Par	*LL*	BIC	aBIC	CIAC	AWE	BLRT	VLMR	BF	*cmP_k*
Class-1	5	−1,294.0	2,618.6	2,602.7	2,623.6	2,664.1	–	–	>100	<.001
Class-2	11	−1,259.2	2,585.6	2,550.7	2,596.6	2,685.7	<.001	<.001	>100	1.00
Class-3	17	−1,249.2	2,602.0	2,548.1	2,619.0	2,756.7	<.001	0.12	>100	<.001
Class-4	23	−1,244.3	2,628.8	2,555.8	2,651.8	2,838.1	0.27	0.01	>100	<.001
Class-5	29	−1,242.4	2,661.7	2,569.7	2,690.7	2,925.6	0.67	0.46	>100	<.001
Class-6	35	−1,240.7	2,695.0	2,583.9	2,730.0	3,013.5	0.67	0.49	–	<.001

[1]Note. Par = parameters; LL = log likelihood; BIC = bayesian information criterion; aBIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE = approximate weight of evidence criterion; BLRT = bootstrapped likelihood ratio test p-value; VLMR = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test p-value; cmPk = approximate correct model probability.

Create Model Fit Summary Table for Latent Transition Analysis (LTA)

STEP 1. Read in models from location where enumeration files are stored for time points 1 & 2 (e.g., `enum_LCA_time1` & `enum_LCA_time1` sub-folders)

```
# TIME 1
output_enum_t1 <- readModels(here("enum_LCA_time1"), quiet = TRUE)

# TIME 2
output_enum_t2 <- readModels(here("enum_LCA_time2"), quiet = TRUE)
```

STEP 2. Extract relevant fit statistics from `mplusObject()` (e.g., named `output_enum` in this example)

```
enum_extract1 <- LatexSummaryTable(output_enum_t1,
                keepCols = c("Title", "Parameters", "LL", "BIC", "aBIC",
                             "BLRT_PValue", "T11_VLMR_PValue","Observations"))

enum_extract2 <- LatexSummaryTable(output_enum_t2,
                keepCols = c("Title", "Parameters", "LL", "BIC", "aBIC",
                             "BLRT_PValue", "T11_VLMR_PValue","Observations"))
```

STEP 3. Calculate additional fit indices derived from the Log Likelihood (LL) as recommended in Nylund et al., 2007.

```
allFit1 <- enum_extract1 %>%
  mutate(aBIC = -2*LL+Parameters*log((Observations+2)/24)) %>%
  mutate(CIAC = -2*LL+Parameters*(log(Observations)+1)) %>%
  mutate(AWE = -2*LL+2*Parameters*(log(Observations)+1.5)) %>%
  mutate(SIC = -.5*BIC) %>%
  mutate(expSIC = exp(SIC - max(SIC))) %>%
  mutate(BF = exp(SIC-lead(SIC))) %>%
  mutate(cmPk = expSIC/sum(expSIC)) %>%
  select(1:5,9:10,6:7,13,14) %>%
  arrange(Parameters)

allFit2 <- enum_extract2 %>%
  mutate(aBIC = -2*LL+Parameters*log((Observations+2)/24)) %>%
  mutate(CIAC = -2*LL+Parameters*(log(Observations)+1)) %>%
  mutate(AWE = -2*LL+2*Parameters*(log(Observations)+1.5)) %>%
  mutate(SIC = -.5*BIC) %>%
  mutate(expSIC = exp(SIC - max(SIC))) %>%
  mutate(BF = exp(SIC-lead(SIC))) %>%
  mutate(cmPk = expSIC/sum(expSIC)) %>%
  select(1:5,9:10,6:7,13,14) %>%
  arrange(Parameters)

allFit <- full_join(allFit1,allFit2)
```

STEP 4. Generate and format the model fit table using the {gt} package.

```
allFit %>%
  mutate(Title = str_remove(Title, "_Time*")) %>%
  gt() %>%
  tab_header(
    title = md("**Model Fit Summary Table**"), subtitle = md(" ")) %>%
  cols_label(
    Title = "Classes",
    Parameters = md("Par"),
    LL = md("*LL*"),
    T11_VLMR_PValue = "VLMR",
    BLRT_PValue = "BLRT",
    BF = md("BF"),
    cmPk = md("*cmPk*")) %>%
  tab_footnote(
    footnote = md(
    "*Note.* Par = Parameters; *LL* = model log likelihood;
      BIC = Bayesian information criterion;
      aBIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion;
      AWE = approximate weight of evidence criterion;
      BLRT = bootstrapped likelihood ratio test p-value;
      VLMR = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test p-value;
      *cmPk* = approximate correct model probability."),
    locations = cells_title()) %>%
  tab_options(column_labels.font.weight = "bold") %>%
  fmt_number(10,decimals = 2,
             drop_trailing_zeros=TRUE,
             suffixing = TRUE) %>%
  fmt_number(c(3:9,11),
             decimals = 1) %>%
  fmt_missing(1:11,
              missing_text = "--") %>%
  fmt(c(8:9,11),
    fns = function(x)
    ifelse(x<0.001, "<.001",
           scales::number(x, accuracy = 0.01))) %>%
  fmt(10,
      fns = function(x)
      ifelse(x>100, ">100",
             scales::number(x, accuracy = .1))) %>%
  tab_row_group(
    group = "Time-1",
    rows = 1:6) %>%
  tab_row_group(
    group = "Time-2",
    rows = 7:12) %>%
  row_group_order(groups = c("Time-1","Time-2"))
```

Model Fit Summary Table[1]

Classes	Par	*LL*	BIC	aBIC	CIAC	AWE	BLRT	VLMR	BF	*cmPk*
Time-1										
C1_LCA1	5	−10,250.6	20,541.3	20,525.5	20,546.3	20,596.5	–	–	–	<.001
C2_LCA1	11	−8,785.3	17,658.9	17,624.0	17,669.9	17,780.2	<.001	<.001	>100	<.001
C3_LCA1	17	−8,693.6	17,523.6	17,469.6	17,540.6	17,711.0	<.001	<.001	>100	0.00
C4_LCA1	23	−8,664.1	17,512.8	17,439.7	17,535.8	17,766.4	<.001	<.001	>100	1.00
C5_LCA1	29	−8,662.4	17,557.5	17,465.4	17,586.5	17,877.3	1.00	0.66	>100	<.001
C6_LCA1	35	−8,661.5	17,604.0	17,492.8	17,639.0	17,989.9	1.00	0.75	>100	<.001
Time-2										
C1_LCA2	5	−7,658.8	15,356.2	15,340.3	15,361.2	15,409.8	–	–	–	<.001
C2_LCA2	11	−6,073.8	12,232.6	12,197.6	12,243.6	12,350.5	<.001	<.001	>100	<.001
C3_LCA2	17	−5,988.4	12,108.0	12,054.0	12,125.0	12,290.3	<.001	<.001	>100	0.32
C4_LCA2	23	−5,964.4	12,106.5	12,033.4	12,129.5	12,353.1	<.001	0.00	2.1	0.68
C5_LCA2	29	−5,961.7	12,147.3	12,055.2	12,176.3	12,458.2	0.31	0.36	>100	<.001
C6_LCA2	35	−5,961.3	12,192.8	12,081.6	12,227.8	12,568.1	1.00	0.50	>100	<.001

[1]Note. Par = Parameters; LL = model log likelihood; BIC = Bayesian information criterion; aBIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE = approximate weight of evidence criterion; BLRT = bootstrapped likelihood ratio test p-value; VLMR = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test p-value; cmPk = approximate correct model probability.

If using this tutorial to produce tables for publication it would be greatly appreciated if you cite this resource using the citation provided here:

Garber, A. C. (2021). Creating Summary Fit Tables for LCA and LTA Analyses Using MplusAutomation. Retrieved from psyarxiv.com/uq2fh

---

References

Hallquist, Michael N., and Joshua F. Wiley. 2018. "MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus." Structural Equation Modeling, 1–18. https://doi.org/10.1080/10705511.2017.1402334.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural equation modeling: A multidisciplinary Journal, 14(4), 535-569.

R Core Team. 2019.R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." Journal of Open Source Software, 4(43), 1686. doi: 10.21105/joss.01686.