IMMERSE Training

June 5-8, 2003

University of California, Santa Barbara

UC SANTA BARBARA

Day 2: Mixture Modeling and Latent Class Analysis (Part 1)

IMMERSE Training

University of California, Santa Barbara

Please do not copy or distribute without permission.

Please do not copy or distribute without permission.

UC SANTA BARBARA

Day 1 Reflections

IMMERSE Project – IES funded Training Grant (R305B220021)

What is one new thing you learned on Day 1 about:

- Latent class measurement models?
- Mixture model estimation?
- Latent class enumeration?
- Latent class interpretation?
- Mplus?
- MplusAutomation?

Was there a nagging question from Day 1 that kept you up last night? If so, what was it?

IMMERSE Project – IES funded Training Grant (R305B220021)

Today’s Agenda – Part 1

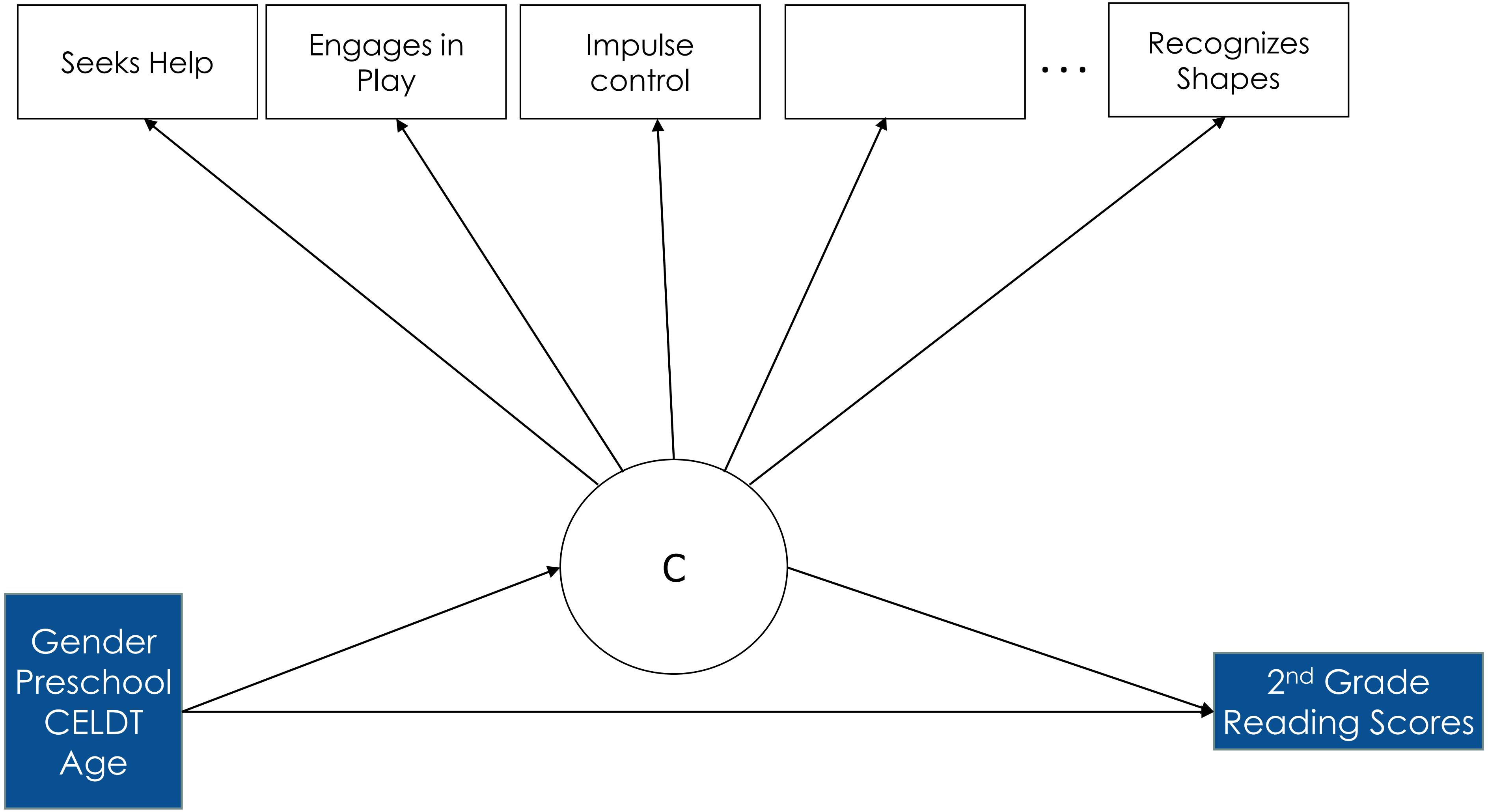
- Overview of adding covariates and distal
- Review of multinomial logistic regression
- Latent class regression
  - Overview
  - ML 3-Step

IMMERSE Project– IES funded Training Grant (R305B220021)

Adding Predictors and Outcomes of Latent Class Membership

Structural model building with one (or more) latent class variables

LCA with Predictors, Covariates, and Outcomes

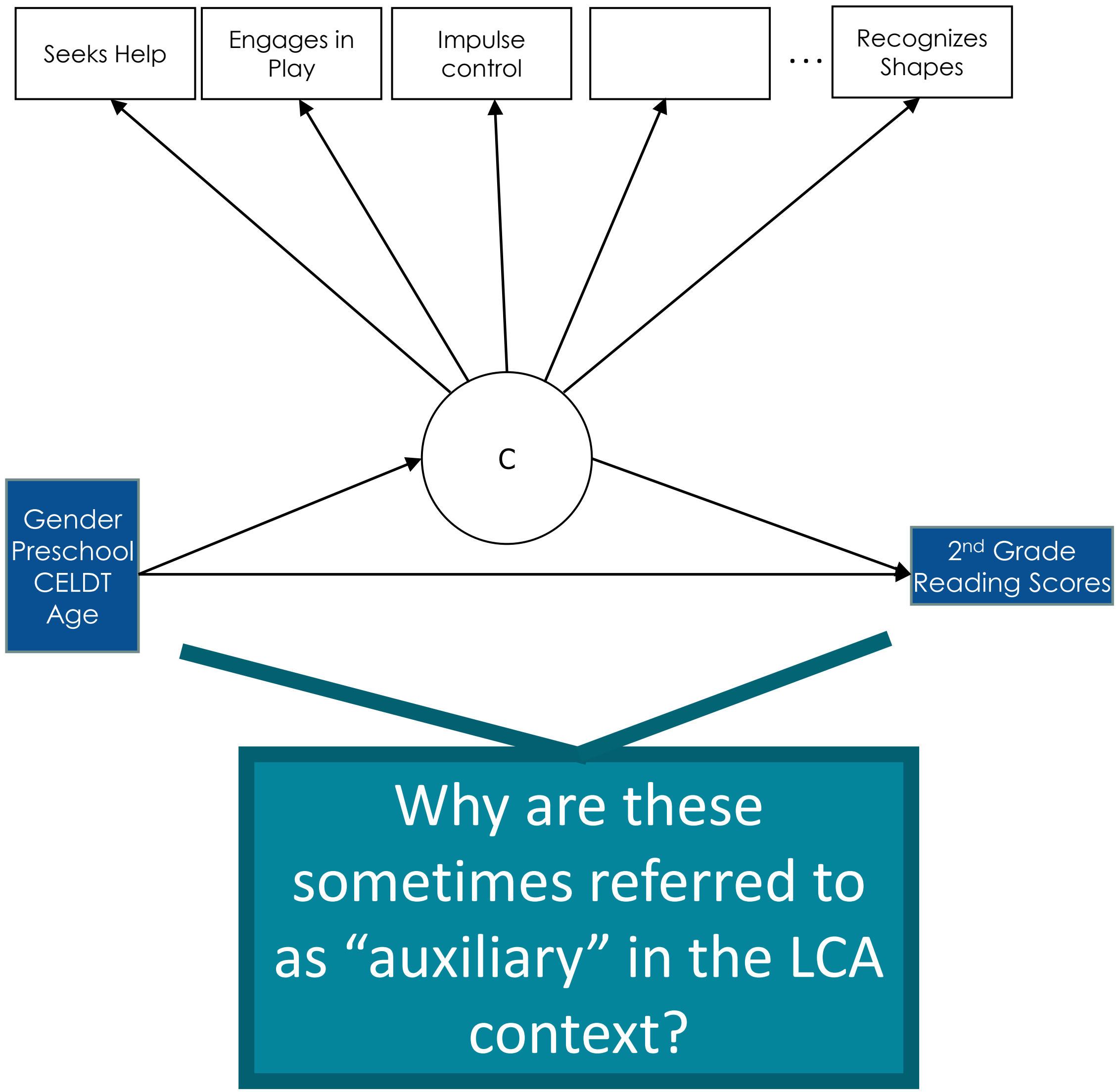


IMMERSE Project– IES funded Training Grant (R305B220021)

"Auxiliary" Variables?

In general, auxiliary variables are variables that _not_ part of the main analysis (i.e., not variables of interest) but assist in the precision or accuracy of the statistical model in a variety of ways. For example:

- Related to missingness (i.e., inclusion needed for MAR to hold in FIML or MI)
- Related to sampling (i.e., used in calculation and/or application of sampling/replicate weights)
- Related to response variable in such a way that it can be used to increase precision of effect estimates



Mixture Model Building Steps

1. Data screening, cleaning, visualizations, and descriptive statistics.

2. Class enumeration process (**without covariates**—except in Auxiliary option of Variable command in Mplus).

-
-
-

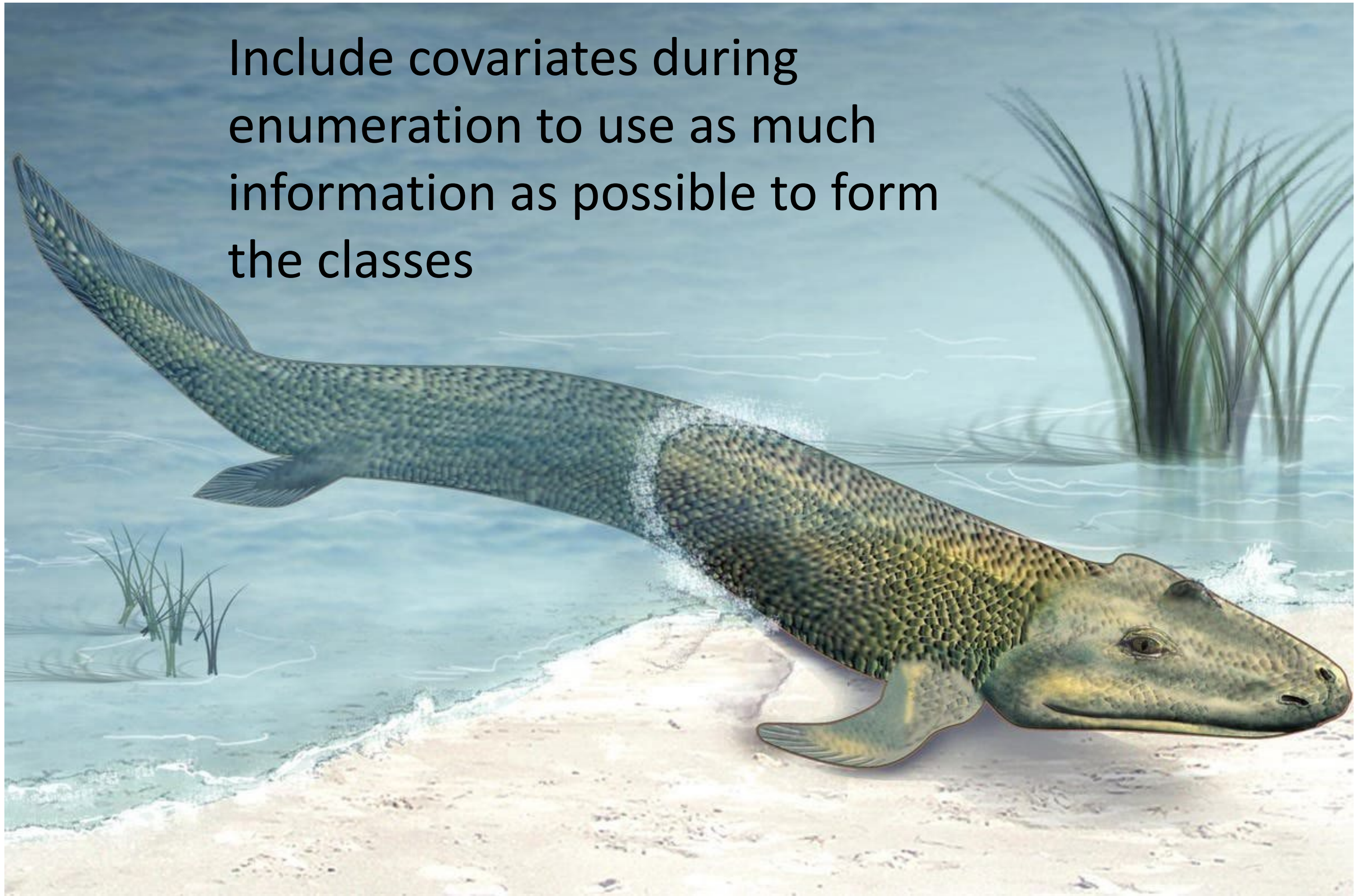
IMMERSE Project– IES funded Training Grant (R305B220021)

Hold the phone!

Have you heard/read recommendations that covariates be included in the class enumeration process because they contain information about class membership that can inform the formation of the classes?

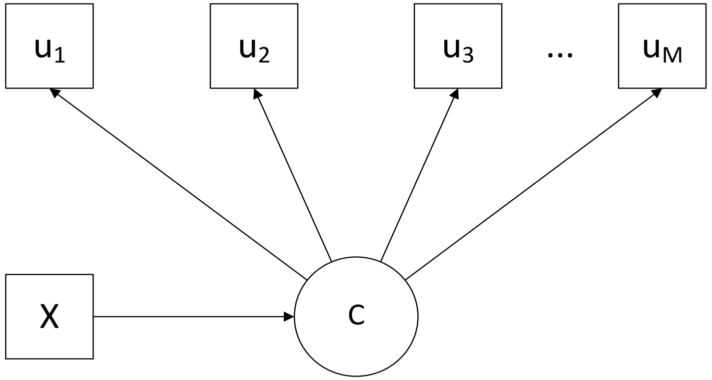
IMMERSE Project– IES funded Training Grant (R305B220021)

Methodological Research Evolving Practice…

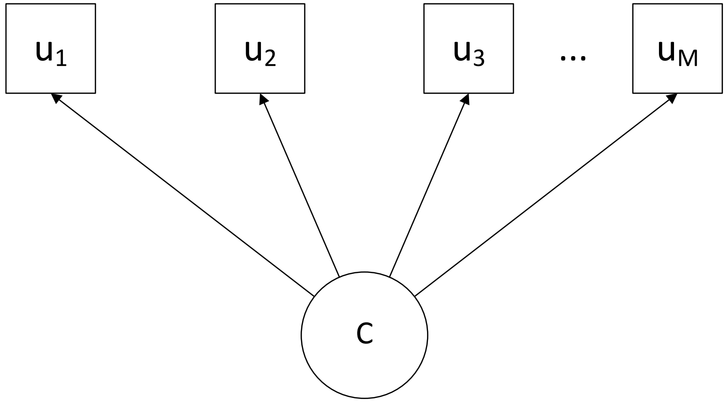


Do not include covariates because formation of the latent classes should not be influenced by or depend on information auxiliary to the measurement model.

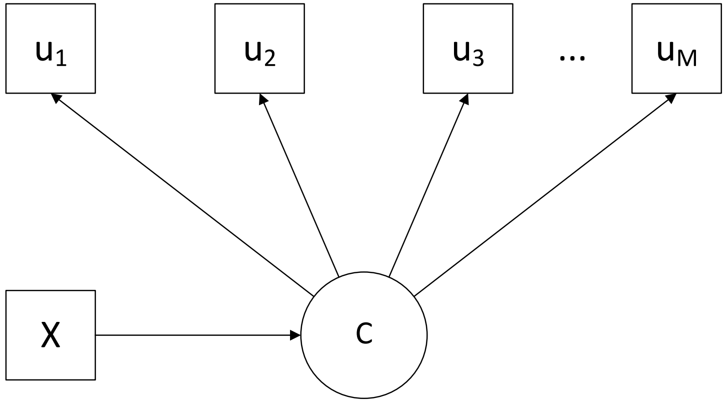
Population Model



Analysis Models



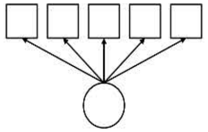
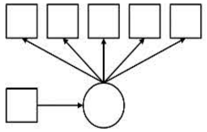
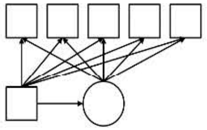
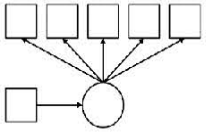
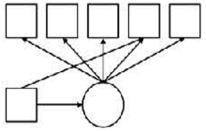
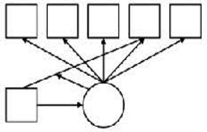
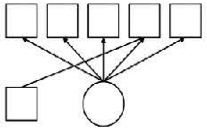
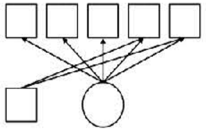
vs.



Nylund-Gibson, K. & Masyn, K. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, DOI: 10.1080/10705511.2016.1221313

IMMERSE Project– IES funded Training Grant (R305B220021)

Average Percent of Replications Selecting 2-class Models as Best According to BIC and BLRT

			Analysis Model			
			M1	M2	M3	M(correct)
Population Model		Fit Criterion*				
PA		BIC	100	100	100	M2
		BLRT	94	88	86	M2
PB		BIC	100	48	100	100
		BLRT	94	2	88	86
PC		BIC	100	33	100	100
		BLRT	93	2	66	60
PD		BIC	100	9	100	100
		BLRT	94	0	90	94
PE		BIC	100	15	100	100
		BLRT	40	0	88	94

* BIC: % of replications for which 2-class model had lowest BIC
BLRT: % for replications for which k=2 class model corresponded to the lowest number of classes for which k vs k+1 had *p*>.05.

Nylund-Gibson, K. & Masyn, K. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, DOI: 10.1080/10705511.2016.1221313

Mixture Model Building Steps

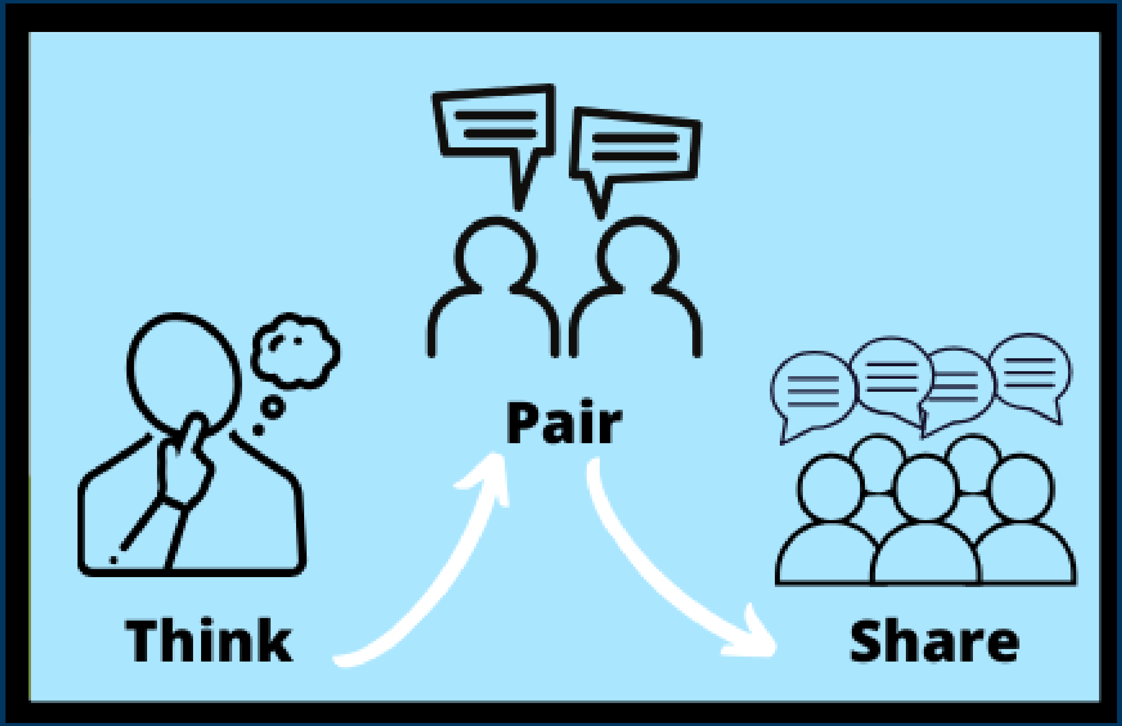
1. Data screening, cleaning, visualizations, and descriptive statistics.
2. Class enumeration process (**without covariates**—except as auxiliary option in Mplus).
3. Select final unconditional model (this is your measurement model).
4. Add hypothesized predictors and covariates (and check for measurement invariance).
5. Add hypothesized (proximal and distal) outcomes.

IMMERSE Project– IES funded Training Grant (R305B220021)

Back up the bus!

We know we're not supposed to do EFA on the same data we use for CFA and SEM, so why is it OK to do a class enumeration and then use the same data for modeling predictors and outcomes of the latent class variable?

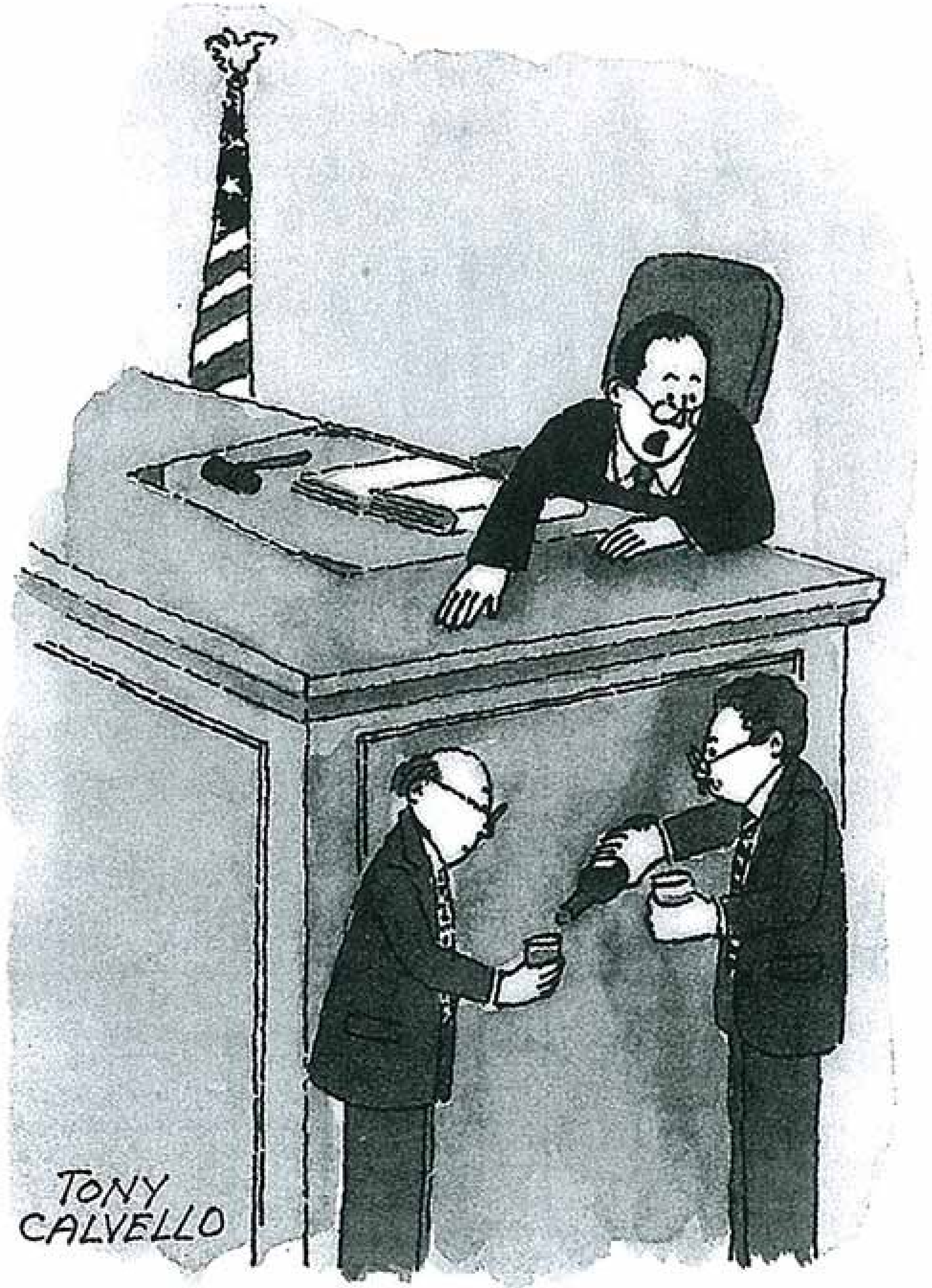
IMMERSE Project– IES funded Training Grant (R305B220021)



- What are your predictor(s)of interest (if any) related to your latent class variable(s)?
- What are your proximal/distal outcomes (if any) related to your latent class variable(s)?
- What might be covariates and control variables you should to include?
- Draw your path diagram.

SIDEBAR:

- Multinomial regression review and Mplus syntax



IMMERSE Project– IES funded Training Grant (R305B220021)

Multinomial Regression

- Multinomial logistic regression is essentially a set of simultaneous binary logistic regressions of the probability in each outcome category versus a reference/baseline category. That is,
  - (j vs. J), ***NOT*** (j vs. ~j)
- For J categories, we have J-1 logit equations, e.g.,
  - 4 categories → 3 binary logistic regressions simultaneously estimated:
    - log odds (1 vs. 4)
    - log odds (2 vs. 4)
    - log odds (3 vs. 4)
    - Note: The odds of (4 vs. 4) is always one and the log odds is always zero.
- Mplus uses the **last** category as the reference/baseline.

We model the following: *Given that the response falls in either category j or J*, what is the log odds that the response is j (instead of J)? That is,

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \qquad \alpha_J = \beta_J = 0$$

This reduces to the familiar binary logistic when J=2.

$$\pi_j = \frac{\exp(\alpha_j + \beta_j x)}{\sum_{h=1}^{J} \left(\exp(\alpha_h + \beta_h x)\right)}, \qquad \alpha_J = \beta_J = 0$$

IMMERSE Project– IES funded Training Grant (R305B220021)

Suppose *J* = 3.

$$\Pr(Y = j) = \pi_j$$

"Odds" are defined as p / (1-p).
How is the Pr(Y = 1) / Pr(Y = 3) an odds?

$$\log\left(\frac{\pi_1}{\pi_3}\right) = \alpha_1 + \beta_1 x$$

$$\log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x$$

$$\log\left(\frac{\pi_3}{\pi_3}\right) = \alpha_3 + \beta_3 x = 0 + 0x = 0$$

$$odds(\text{A}) = \frac{\Pr(\text{A})}{1-\Pr(\text{A})} = \frac{\Pr(\text{A})}{\Pr(\text{not A})}$$

$$\frac{\Pr(Y=1)}{\Pr(Y=3)} \neq odds(Y=1) = \frac{\Pr(Y=1)}{1-\Pr(Y=1)} = \frac{\Pr(Y=1)}{\Pr(Y=2 \text{ or } Y=3)}$$

$$\frac{\Pr(Y=1)}{\Pr(Y=3)} = odds(Y=1 \mid Y=1 \text{ or } Y=3)$$

IMMERSE Project– IES funded Training Grant (R305B220021)

$$\pi_j = \frac{\exp(\alpha_j + \beta_j x)}{\sum_{h=1}^{J} \left(\exp(\alpha_h + \beta_h x) \right)}, \qquad \alpha_J = \beta_J = 0$$

$$\sum_{h=1}^{J} \left(\exp(\alpha_h + \beta_h x) \right) = \sum_{h=1}^{3} \left(\exp(\alpha_h + \beta_h x) \right)$$

$$= \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x) + \exp(\alpha_3 + \beta_3 x)$$

$$= \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x) + 1$$

$$\exp\left(\alpha_3 + \beta_3 x\right)$$
$$= \exp\left(0 + 0x\right)$$
$$= \exp\left(0\right)$$
$$= 1$$

IMMERSE Project– IES funded Training Grant (R305B220021)

$$\Pr(Y=1)=\frac{\exp(\alpha_1+\beta_1 x)}{\exp(\alpha_1+\beta_1 x)+\exp(\alpha_2+\beta_2 x)+1}$$

$$\Pr(Y=2)=\frac{\exp(\alpha_2+\beta_2 x)}{\exp(\alpha_1+\beta_1 x)+\exp(\alpha_2+\beta_2 x)+1}$$

$$\Pr(Y=3)=\frac{1}{\exp(\alpha_1+\beta_1 x)+\exp(\alpha_2+\beta_2 x)+1}$$

$$\Pr(Y=1)+\Pr(Y=2)+\Pr(Y=3)=1$$

Notice:

- There are three probabilities.
- There are three terms being summed in the denominator.
- Each term appears in the numerator of one probability.
- The denominator is the same for all three.

IMMERSE Project– IES funded Training Grant (R305B220021)

Interpreting Estimates

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \beta_j x, \qquad \alpha_J = \beta_J = 0$$

$$\alpha_j = \log\left(odds\left(Y = j \mid \left(Y = j \text{ or } Y = J\right) \text{ and } X = 0\right)\right)$$

IMMERSE Project– IES funded Training Grant (R305B220021)

Interpreting EXP(β_j)

- Conditional Odds Ratio (COR)
  - OR for being in category j versus J (given membership in either j or J) corresponding to a positive one-unit difference in X.
- Relative Risk Ratio (RRR)
  - Ratio of the RR for category j corresponding to a positive one-unit difference in X to the RR for category J corresponding to a positive one-unit difference in X.

IMMERSE Project– IES funded Training Grant (R305B220021)

Nominal *Dependent* Variables in Mplus

VARIABLE: !Mplus command

Nominal are **names of unordered categorical *dependent* variables (multinomial);**

- The intercepts and slopes for each logit equation are referred to in the MODEL command by adding to the variable name the pound sign (#) followed by a number. For example,
  - The two intercepts for a three-category nominal variable, u, are referred to as "[u#1]" and "[u#2]".
  - The two slopes for a predictor, x, are "u#1 on x" and "u#2 on x"
  - Note: If you specify u2 as a nominal endogenous variable in the variable command and then write "u on x" in the model command, Mplus will automatically expand that internally to a multinomial logistic regression with two intercepts and two slopes.

Multinomial Regression Example

Camera Marketing Study

Sample of 735 of individuals surveyed by a market research group for the purposes of investigating the role of age and "gender" (outdated binary—this is old data) in digital camera brand choices. Variables for the study include

- brand
  - 1 = Canon
  - 2 = Kodak
  - 3 = Nikon
- female
  - 1 = female
  - 0 = male
- age (in years)

Data Snapshot

		brand			Total
		1 canon	2 kodak	3 nikon	
female	1	115	208	143	466
	0	92	99	78	269
Total		207	307	221	735

Age: Min = 24 yrs | Max = 38 yrs | Mean = 32.9 yrs | SD = 2.3 yrs

IMMERSE Project– IES funded Training Grant (R305B220021)

Mplus Input: Multinomial regression of *brand* on *female*

```
DATA:
File is camera.dat;
VARIABLE:
Names are brand female age;
!Brand: 1 = Canon, 2 = Kodak, 3 = Nikon
UseVariables are brand female;

Nominal are brand;

ANALYSIS:
Estimator = MLR;

MODEL:
brand on female;

OUTPUT:
svalues;
```

Equivalent MODEL statements:

- `brand#1 brand#2 on female;`
- `brand#1 on female;`
  `brand#2 on female;`

(Select) Mplus Output

```
MODEL FIT INFORMATION

Number of Free Parameters                        4

Loglikelihood

          H0 Value                        -791.861
          H0 Scaling Correction Factor      1.0000
            for MLR

Information Criteria

          Akaike (AIC)                    1591.723
          Bayesian (BIC)                  1610.122
          Sample-Size Adjusted BIC        1597.421
            (n* = (n + 2) / 24)
```

What are the four parameters being estimated?

IMMERSE Project– IES funded Training Grant (R305B220021)

(Select) Mplus Output

```
MODEL RESULTS

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

 BRAND#1    ON
    FEMALE            -0.383      0.198     -1.930      0.054

 BRAND#2    ON
    FEMALE             0.136      0.186      0.731      0.465

 Intercepts
    BRAND#1            0.165      0.154      1.073      0.283
    BRAND#2            0.238      0.151      1.575      0.115
```

(Select) Mplus Output

Why 95% CI instead of
Est./S.E. P-Value ?

LOGISTIC REGRESSION ODDS RATIO RESULTS

	Estimate	S.E.	95% C.I. Lower 2.5%	Upper 2.5%
BRAND#1 ON				
FEMALE	0.682	0.135	0.462	1.006
BRAND#2 ON				
FEMALE	1.146	0.214	0.795	1.651

What is the interpretation of this OR?

(Select) Mplus Output

```
MODEL RESULTS

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

 BRAND#1    ON
    FEMALE            -0.383      0.198     -1.930      0.054

 BRAND#2    ON
    FEMALE             0.136      0.186      0.731      0.465
```

Overall, is there evidence that sex is associated with camera brand choice?

IMMERSE Project– IES funded Training Grant (R305B220021)

(Select) Mplus Output

```
MODEL COMMAND WITH FINAL ESTIMATES USED AS STARTING VALUES

     brand#1 ON female*-0.38299;
     brand#2 ON female*0.13628;

     [ brand#1*0.16508 ];
     [ brand#2*0.23841 ];
```

Produced by "OUTPUT: Svalues;" One line of syntax for each parameter—in this case, four—with start values set the final MLEs from the Model Results in the same output.

Mplus Input w/ Omnibus Test

```
.
.
.
MODEL:
!MODEL COMMAND WITH FINAL ESTIMATES USED AS STARTING VALUES

     brand#1 ON female*-0.38299 (femCvN);
     brand#2 ON female*0.13628 (femKvN);

     [ brand#1*0.16508 ] (intCvN);
     [ brand#2*0.23841 ] (intKvN);

Model Test:
  0 = femCvN;
  0 = femKvN;
```

A user-inputted *start* value follows an "*". A user-specified *fixed* value follow an "@". A parameter label is given in parentheses before ";".

What is this testing? Null hypothesis? Alternative hypothesis?

IMMERSE Project– IES funded Training Grant (R305B220021)

(Select) Mplus Output

```
Number of Free Parameters                         4

Loglikelihood
          H0 Value                         -791.861
.
.
.
Wald Test of Parameter Constraints

          Value                               8.097
          Degrees of Freedom                      2
          P-Value                            0.0174
```

This multivariate Wald test of parameter constraints is asymptotically equivalent to the likelihood ratio (chi-square) test of nested model comparing this model (full) to the constrained/nested model with MODEL:
brand#1 on female @0;
brand #2 on female@0;

What is the statistical inference based on this test result (using $\alpha = .05$)?

IMMERSE Project– IES funded Training Grant (R305B220021)

Mplus Input w/ Alternate COR/RRR

```
.
.
.
MODEL:
!MODEL COMMAND WITH FINAL ESTIMATES USED AS STARTING VALUES

     brand#1 ON female*-0.38299 (femCvN);
     brand#2 ON female*0.13628 (femKvN);

     [ brand#1*0.16508 ] (intCvN);
     [ brand#2*0.23841 ] (intKvN);

Model Constraint:
  New(femCvK efemCvK);
  femCvK = femCvN - femKvN;
  efemCvK =exp(femCvK);
```


IMMERSE Project– IES funded Training Grant (R305B220021)

(Select) Mplus Output

```
Estimate          S.E.  Est./S.E.     P-Value
 BRAND#1      ON
     FEMALE              -0.383      0.198     -1.930      0.054

 BRAND#2      ON
     FEMALE               0.136      0.186      0.731      0.465

 Intercepts
     BRAND#1              0.165      0.154      1.073      0.283
     BRAND#2              0.238      0.151      1.575      0.115

New/Additional Parameters
     FEMCVK              -0.519      0.186     -2.797      0.005
     EFEMCVK              0.595      0.110      5.386      0.000
```

What is the interpretation of this OR?

IMMERSE Project– IES funded Training Grant (R305B220021)

Mplus Input: Multinomial regression of *brand* on *female* & *age*

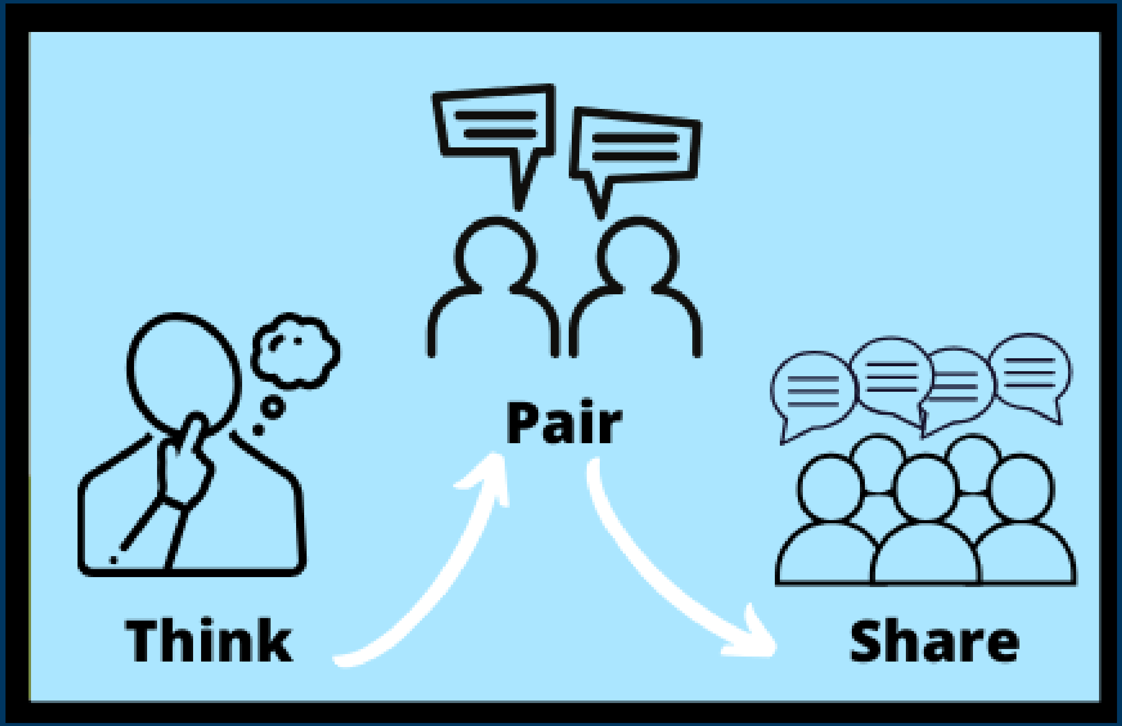
```
DATA:
File is camera.dat;
VARIABLE:
Names are brand female age;
!Brand: 1 = Canon, 2 = Kodak, 3 = Nikon
UseVariables are brand female age;

Nominal are brand;

ANALYSIS:
Estimator = MLR;

MODEL:
brand on female age;

OUTPUT:
svalues;
```


IMMERSE Project– IES funded Training Grant (R305B220021)

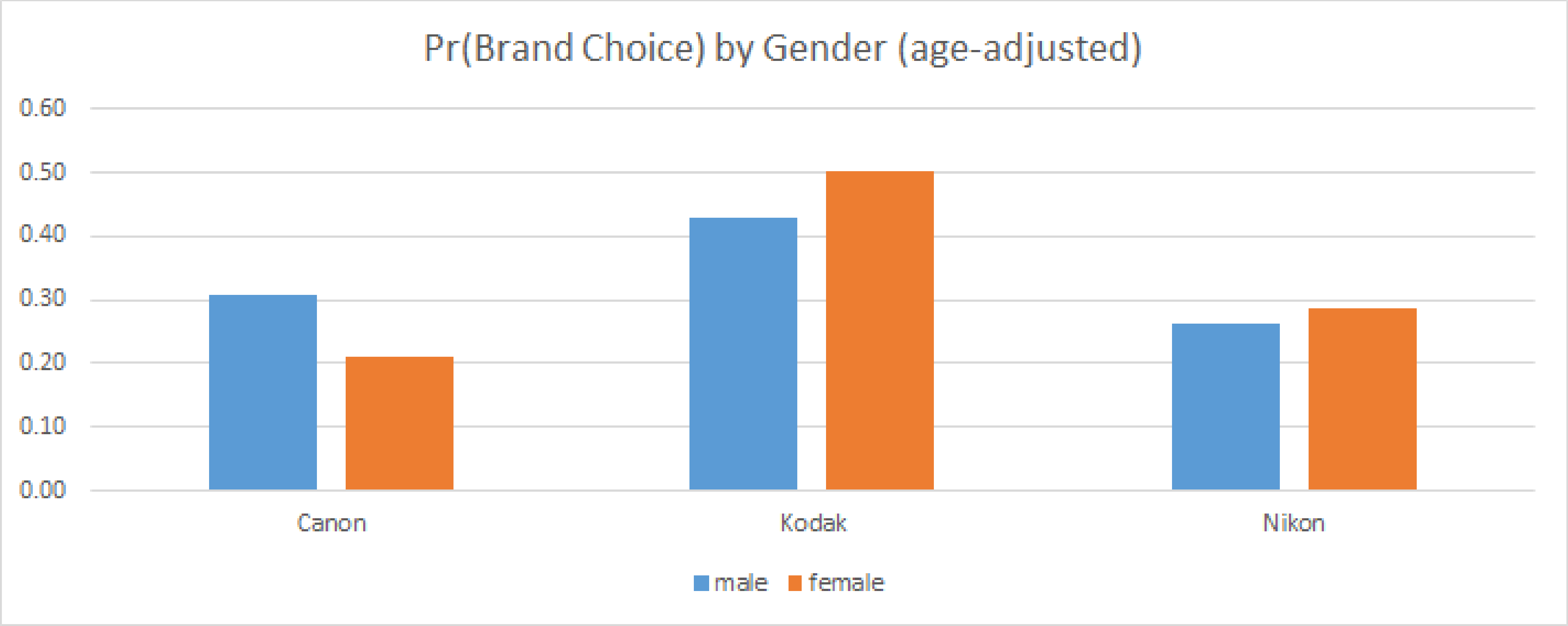
(Select) Mplus Output

```
MODEL RESULTS

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

 BRAND#1    ON
    FEMALE            -0.466      0.227     -2.057      0.040
    AGE               -0.686      0.072     -9.497      0.000

 BRAND#2    ON
    FEMALE             0.058      0.196      0.296      0.768
    AGE               -0.318      0.046     -6.882      0.000

 Intercepts
    BRAND#1           22.721      2.378      9.554      0.000
    BRAND#2           10.947      1.571      6.969      0.000
```

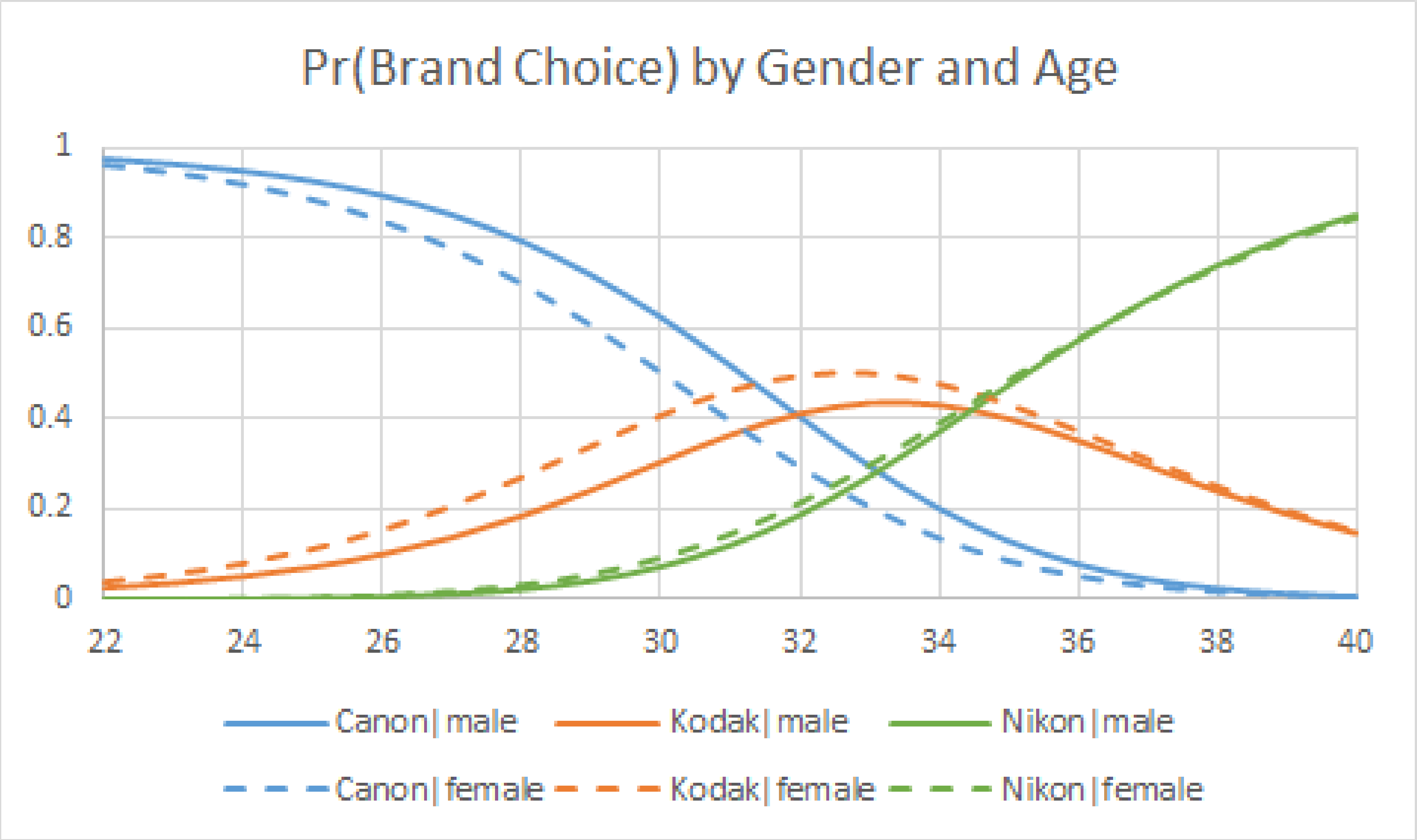

IMMERSE Project– IES funded Training Grant (R305B220021)



- How would you test the effect of sex on camera brand choice adjusted for age?
- How would you test the effect of age on camera brand choice adjusted for sex?
- How would you test for an interaction effect between age and sex on camera brand choice?
- Based on the model results, which camera brand is the most popular amongst females, adjusted for age? What about for males?
- What matters more for camera brand choice, age or sex?
- How could you depict the adjusted effects of sex and age on camera brand choice in the same graph?



IMMERSE Project– IES funded Training Grant (R305B220021)

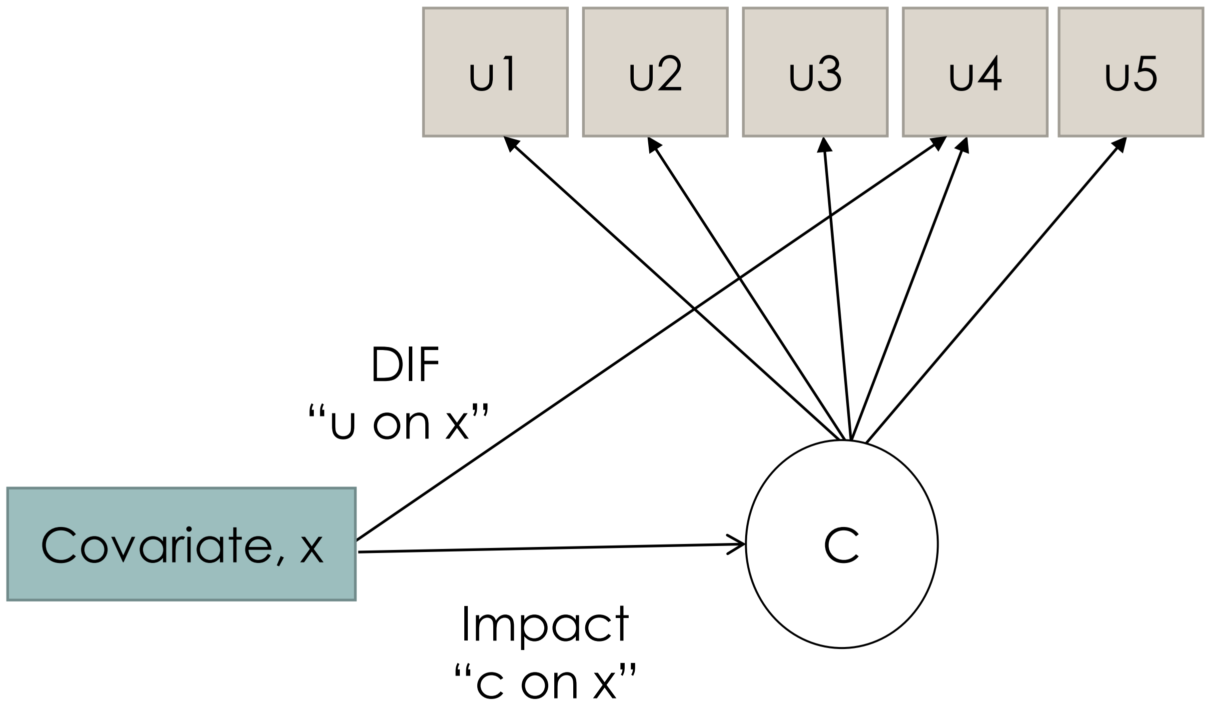


IMMERSE Project– IES funded Training Grant (R305B220021)

UC SANTA BARBARA

Latent Class Regression

Covariates and Mixture Models (LC-MIMIC)



IMMERSE Project– IES funded Training Grant (R305B220021)

Latent Class Regression

- Categorical latent variable
- Continuous or categorical covariates with effects on indicators (u, y) and/or effects on the latent class variable
  - Effects on items can also be thought of as differential item functioning (DIF), i.e., measurement non-invariance.
    - These are sometimes referred to as “direct effects” because there is a direct path from an observed exogenous variables to an observed endogenous variable.
  - Variables with effects on latent class variable can also be thought of as predictors or correlates of class membership.
    - These are sometimes referred to as “impact effects” because the covariate is directly impacting the latent construct.
    - These are sometimes referred to as “indirect effects” because there is an indirect path from the exogenous variable to the observed items via the latent class variable (in comparison to the DIF effects).

IMMERSE Project– IES funded Training Grant (R305B220021)

"C on X" = Multinomial Regression

- Mplus uses the **last** category/class as the baseline.
- So for K classes, we have *K-1* logit equations.
- We model the following: *Given membership in either Class k or K*, what is the log odds that class membership is k (instead of K)? That is,

$$\log\left(\frac{\Pr(c=k \mid c=k \text{ or } K)}{\Pr(c=K \mid c=k \text{ or } K)}\right)=\alpha_k+\gamma_k x,$$

$$\alpha_K=\gamma_K=0$$

Modeling Steps with Predictors and Covariates

1. Data cleaning/screening
2. Unconditional mixture model → finalize measurement model
3. Explore DIF with predictor and covariate variables
4. Test predictor effects (impact) on latent class variable
   a) Omnibus test of association with latent class variable
   b) Pairwise differences (if omnibus test indicates an association)

IMMERSE Project– IES funded Training Grant (R305B220021)

Stepwise Procedures: A Motivating Example

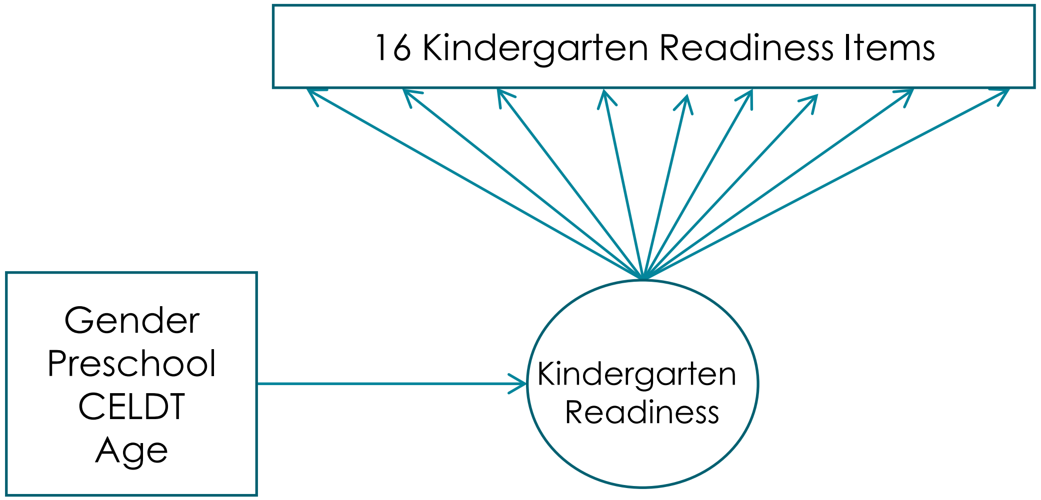
KSEP Example

- Review example of Kindergarten readiness
- Motivate why we need multi-step processes

Note: The specification of covariates in mixture is an active area of research. The current "best practice" is that we use methods called "stepwise" approaches. The currently dominant stepwise approaches involve *three* steps and so we will often refer to stepwise approaches as "three-step" methods.

IMMERSE Project– IES funded Training Grant (R305B220021)

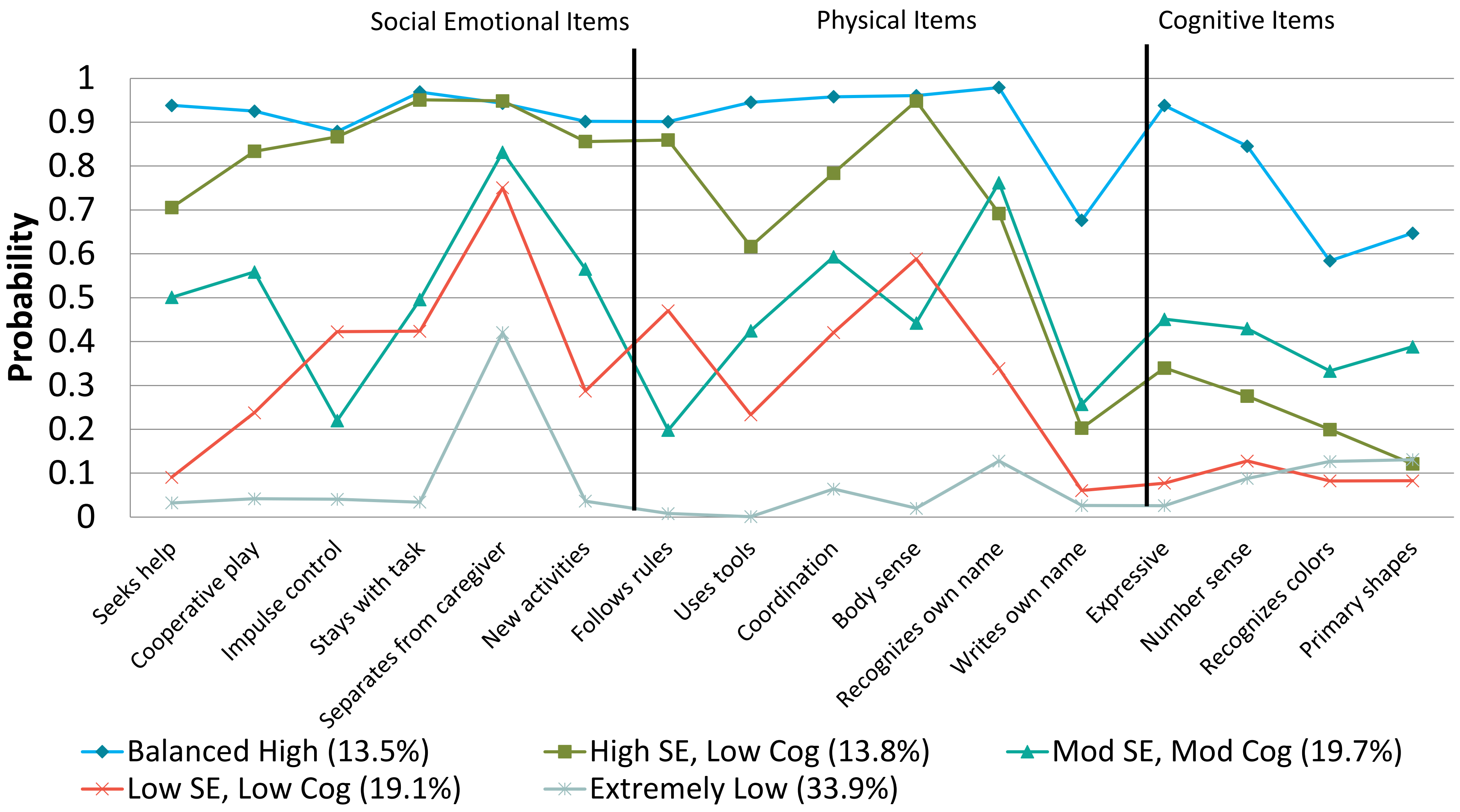
Profiles of Kindergarten Readiness (LCA)



Kindergarten Student Entrance Profile (KSEP)

A. Seeks adult help when appropriate
B. Engages in cooperative play activities with peers
C. Exhibits impulse control and self-regulation
D. Stays with or repeats a task
E. Separates appropriately from caregiver
F. Is enthusiastic and curious in approaching new activities
G. Follows rules when participating in routine activities
H. Uses tools with increasing precision
I. Demonstrates general coordination
J. Demonstrates sense of own body in relation to others
K. Recognizes own written name
L. Writes own name
M. Demonstrates expressive abilities
N. Understands that numbers represent quantity
O. Recognizes colors
P. Recognizes primary shapes

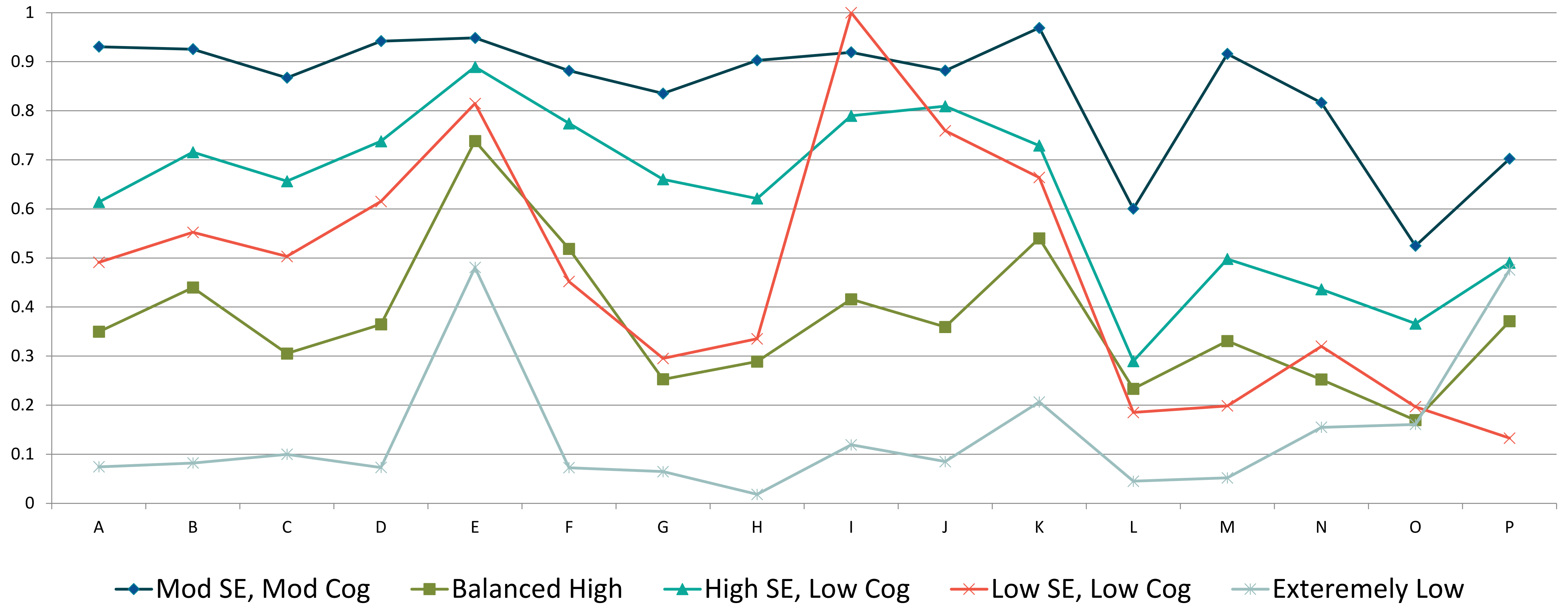
Unconditional 5-class LCA Solution



Quirk, M., Nylund-Gibson, K., & Furlong, M. (2013). Exploring Patterns of Latino/a Children's School Readiness at Kindergarten Entry and their Relations with Grade 2 Achievement. *Early Childhood Research Quarterly, 28(2), 437-449.*

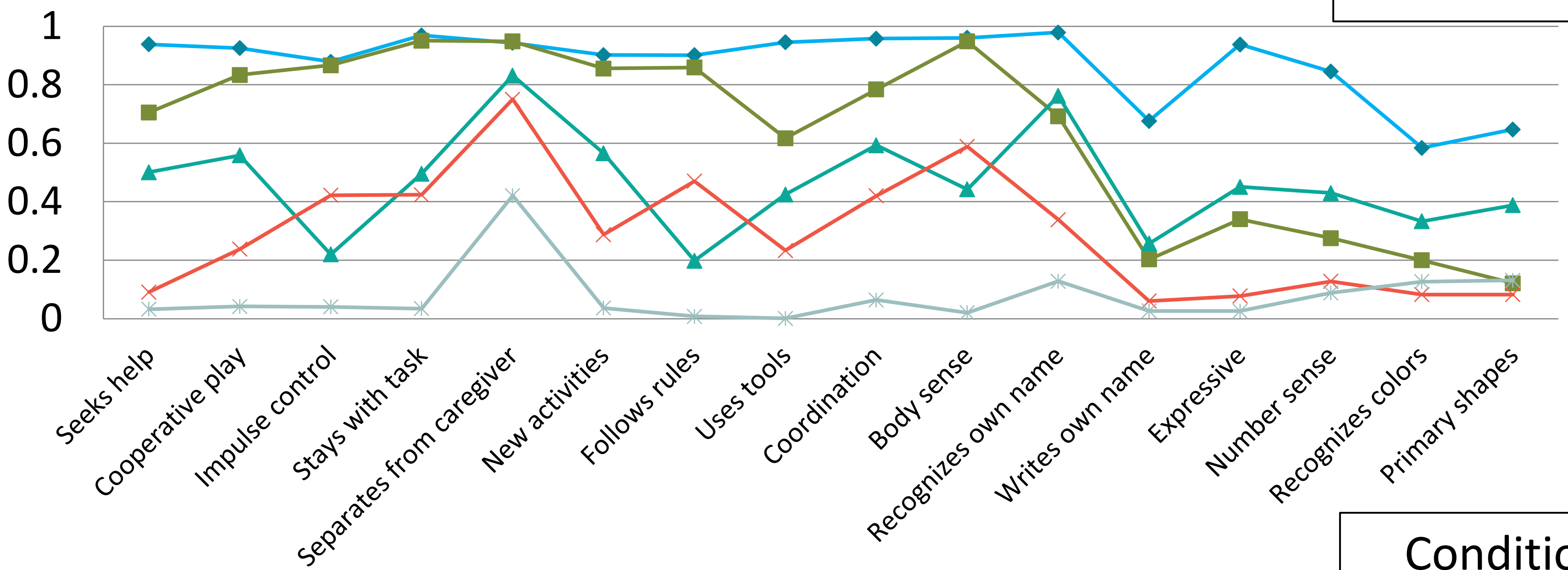
IMMERSE Project– IES funded Training Grant (R305B220021)

LCA with Covariates (conditional; c on x)



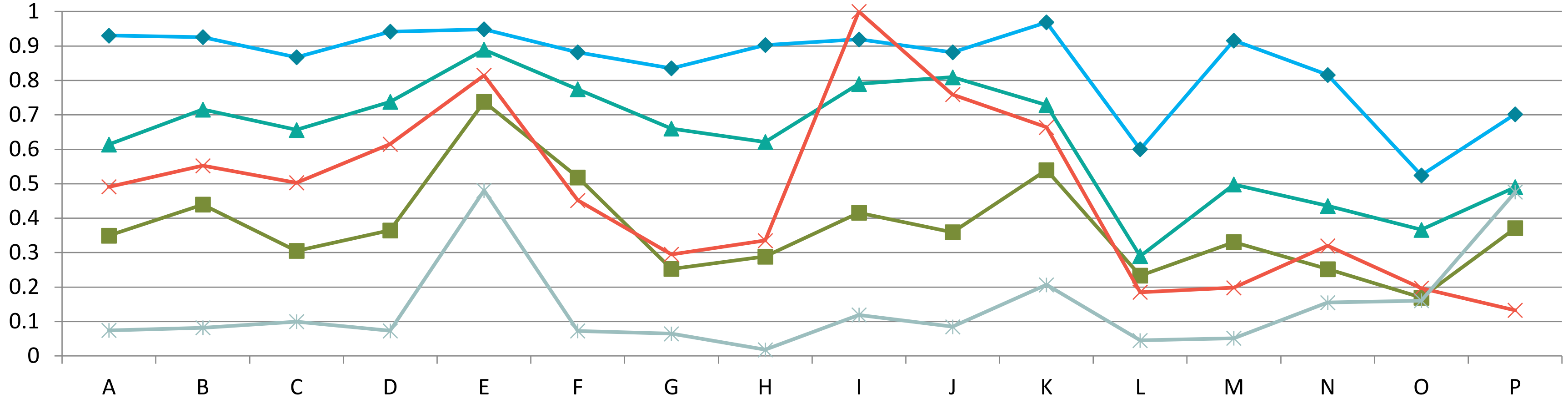
IMMERSE Project– IES funded Training Grant (R305B220021)

Unconditional



What differences do you notice in the 5-class solutions?

Conditional

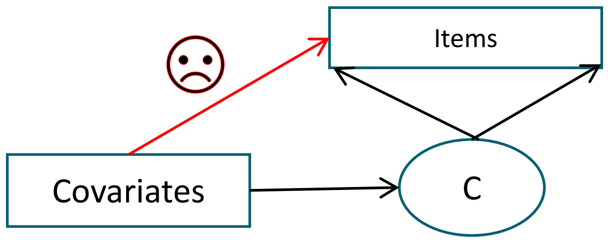


IMMERSE Project– IES funded Training Grant (R305B220021)

UC SANTA BARBARA

Differences in unconditional and conditional models

- Indicates some relationship between indicators and auxiliary variables (covariates) included in the model, i.e., **DIF!**
- Number and type of classes can change
- Not necessarily intended since we want latent classes to be identified solely by indicators
- Uneasy findings by substantive researchers



Note: Stepwise procedures were originally developed to get around the impact of DIF on the latent classes by uncoupling the estimation of the measurement model from the structural model. However, we now know that bias from unaccounted-for DIF in the measurement model can still bias structural effect estimated, even using the stepwise approaches.

Bottom line: There's no getting around or ignoring DIF.

Advance Access publication September 23, 2010

Political Analysis (2010) 18:450–469
doi:10.1093/pan/mpq025

Latent Class Modeling with Covariates: Two Improved Three-Step Approaches

Jeroen K. Vermunt
Department of Methodology and Statistics, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: j.k.vermunt@uvt.nl

However, the one-step approach has certain disadvantages. The first is that it may sometimes be impractical, especially when the number of potential covariates is large, as will typically be the case in a more exploratory study. Each time that a covariate is added or removed not only the prediction model but also the measurement model needs to be reestimated. A second disadvantage is that it introduces additional model building problems, such as whether one should decide about the number of classes in a model with or without covariates. Third, the simultaneous approach does not fit with the logic of most applied researchers, who view introducing covariates as a step that comes after the classification model has been built. Fourth, it assumes that the classification model is built in the same stage of a study as the model used to predict the class membership, which is not necessarily the case. It can even be that the researcher who constructs the typology using an LC model is not the same as the one who uses the typology in a next stage of the study.

Stepwise Procedures: Overview

IMMERSE Project – IES funded Training Grant (R305B220021)

Review of Past Recommended Auxiliary Variable recommendations

1. 1-step
2. Classify-analyze
3. Pseudo-class draws
   - Auxiliary = z (E);
4. ML 3-step
   - Auxiliary = z (DU3step) or (DE3step)
   - Manual 3-step
5. "New" Bayes' Theorem approach by Lanza et al. (2014)
   - Auxiliary = z (DCON) or (DCAT)
6. BCH method 3-step
   - Auxiliary = z (bch)
   - Manual 3-step
7. New 2-step (Bakk & Zuba, 2017)

Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural equation modeling: A multidisciplinary Journal*, *26*(6), 967-985.

IMMERSE Project– IES funded Training Grant (R305B220021)

Review of Past Recommended Auxiliary Variable recommendations

1. ~~1-step~~
2. Classify-analyze
3. ~~Pseudo-class draws~~
   - ~~Auxiliary = z (E);~~
4. ML 3-step
   - Auxiliary = z (DU3step) or (DE3step)
   - Manual 3-step
5. ~~New Bayes' Theorem approach by Lanza et al. (2014)~~
   - ~~Auxiliary = z (DCON) or (DCAT)~~
6. BCH method 3-step
   - Auxiliary = z (bch)
   - Manual 3-step
7. New 2-step (Bakk & Zuba, 2017)

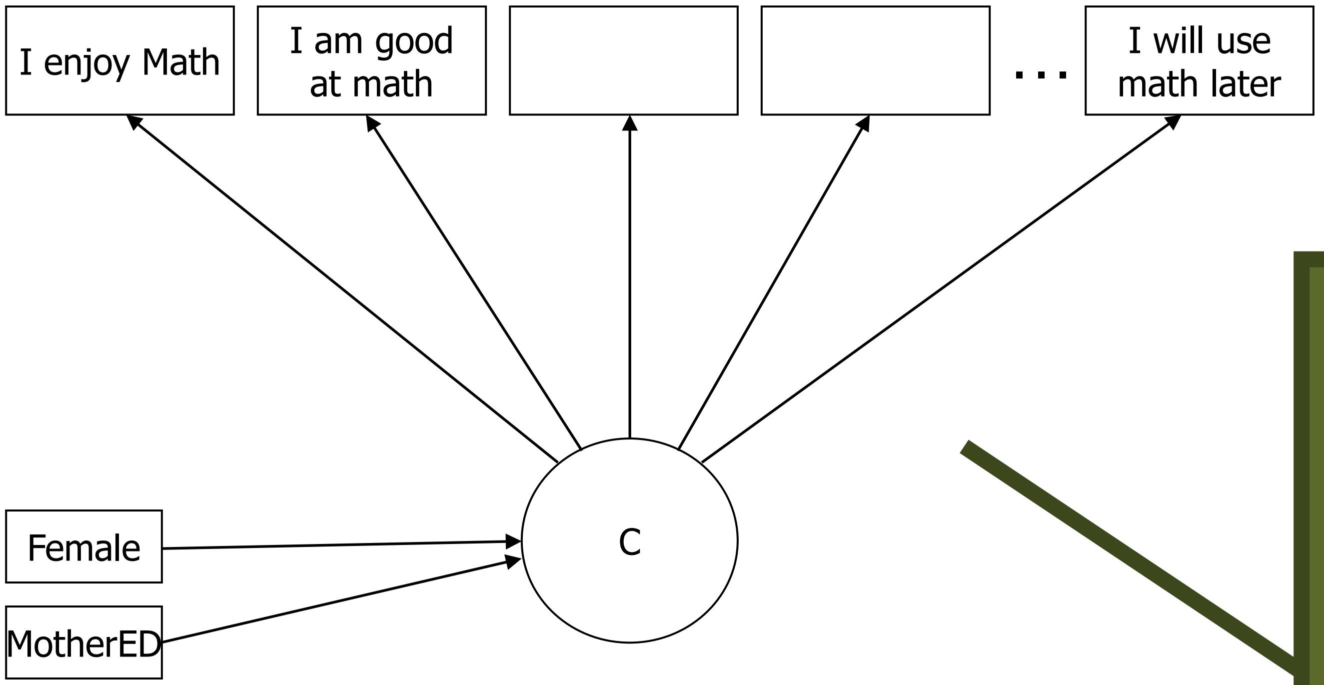
These were recommended as "best practices" at some point.

Note- Mplus still allows for the specification of these approaches.

Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural equation modeling: A multidisciplinary Journal*, *26*(6), 967-985.

IMMERSE Project– IES funded Training Grant (R305B220021)

1-Step: Latent Class regression with two covariates



Measurement model re-estimated simultaneously with covariate effects on latent class variables. Nothing wrong with this—it's what we do in SEM—*but* cannot ignore or not consider covariates as potential source of DIF.

"Old" 2-step: Classify-analyze approach

- Assign individuals to modal class and do subsequent analysis treating class assignment as an observed grouping variable.
- Analogous to predicted values in regression or factor scores in factor analysis
- Doesn't account for uncertainty (i.e., measurement/classification error) in class assignment, which leads to
  - Biased estimates and standard errors
  - Distorted substantive findings

IMMERSE Project– IES funded Training Grant (R305B220021)

Overview

- **Three-Step method**
  - Automatic– done within Mplus
    - Only allows for either only covariates or only distals (not both covariates and distals)
  - Manual- specified by user
    - Allows for both covariates and distal outcomes and their direct relationship
    - Allows for multiple latent class variables (e.g., latent transition analysis)
- **BCH method**
  - Automatic– done within Mplus
    - Only allows for distals
  - Manual
    - Allows for both covariates and distal outcomes and their direct relationship
    - Cannot use with more than one latent class variable

IMMERSE Project– IES funded Training Grant (R305B220021)

Overview

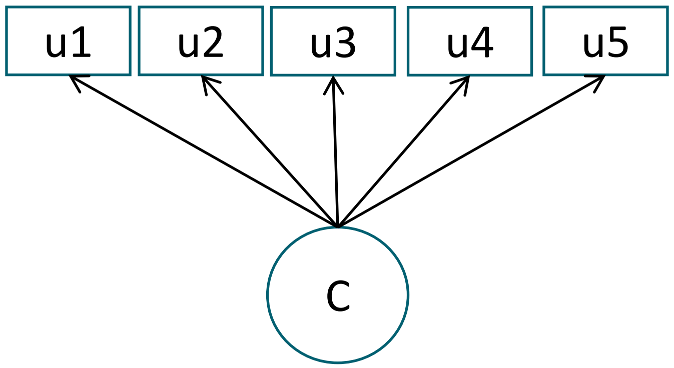
- **Three-Step method**
  - ~~Automatic– done within Mplus~~
    - ~~Only allows for either only covariates or only distals (not both covariates and distals)~~
  - Manual- specified by user
    - Allows for both covariates and distal outcomes and their direct relationship
    - Allows for multiple latent class variables (e.g., latent transition analysis)
- **BCH method**
  - ~~Automatic– done within Mplus~~
    - ~~Only allows for distals~~
  - Manual
    - Allows for both covariates and distal outcomes and their direct relationship
    - Cannot use with more than one latent class variable

IMMERSE Project– IES funded Training Grant (R305B220021)

Manual ML 3-Step

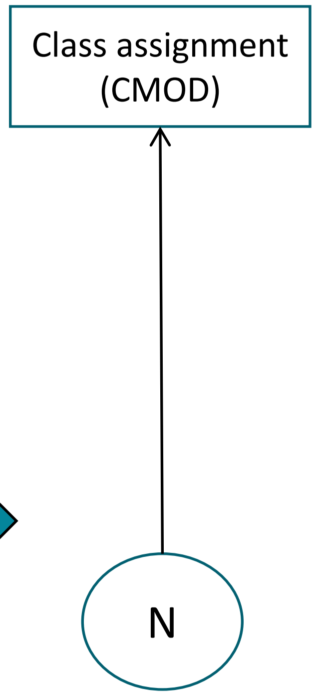
ML 3-Step Overview

Step 1: Estimate the unconditional LCA



Step 2: Assign individuals to their most likely class. Create a new variable with the modal class assignments.

Step 3: In new model, use modal class assignment as single nominal indicator of latent class membership, fixing measurement parameters to reflect the classification error rates estimated in model from (1). "N" has the same number of classes as "C".



ML 3-Step: Incorporating classification error

- ML 3-step accounts for error in the modal class assignment variable (cmod) by including information about misclassification.
- These errors are summarized by averaging errors over individuals for each class. Thus, the error is at the class level.

Classification Probabilities for the M
by Latent Class (Row)

	1	2	3
1	0.894	0.106	0.000
2	0.016	0.925	0.059
3	0.000	0.140	0.860

Logits for the Classification Probabi
by Latent Class (Row)

	1	2	3
1	13.704	11.568	0.000
2	-1.290	2.752	0.000
3	-11.543	-1.819	0.000

IMMERSE Project– IES funded Training Grant (R305B220021)

ML 3-step: Posterior Probabilities

- The LCA model provides information about how well people are classified into each of the latent classes, posterior probabilities.

Classification Probabilities for the Most Likely Latent Class Membership (Column) by Latent Class (Row)

	1	2	3
1	0.894	0.106	0.000
2	0.016	0.925	0.059
3	0.000	0.140	0.860

Average class prob for C1 | assigned to C1

Average class prob for C1 | assigned to C2

- We use this information from the table to specify measurement error in the ML 3-step method.

IMMERSE Project– IES funded Training Grant (R305B220021)

ML 3-step: Posterior Probabilities

Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)

	1	2	3
1	0.894	0.106	0.000
2	0.016	0.925	0.059
3	0.000	0.140	0.860

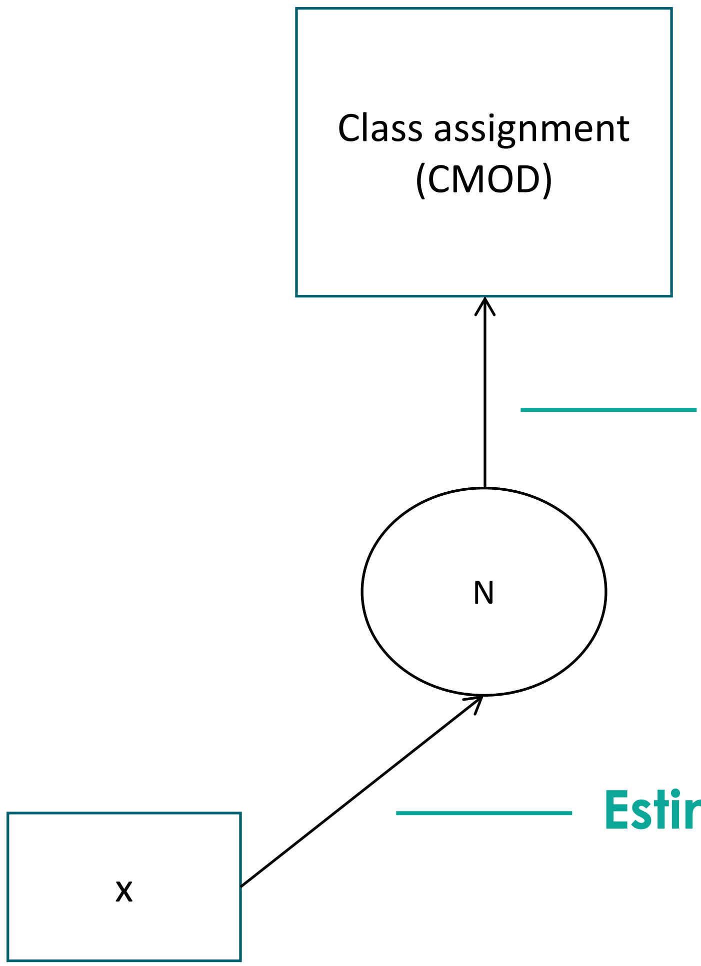
Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)

	1	2	3
1	13.704	11.568	0.000
2	-1.290	2.752	0.000
3	-11.543	-1.819	0.000

Mplus provides you with the logit values

$$q_{c_1,c_2} = P(S = c_1 | C = c_2) = \frac{p_{c_1,c_2} N_{c_1}}{\sum_c p_{c,c_2} N_c} \quad (1)$$

This is how you hand calculate them



Fixed parameters (accounting for error)

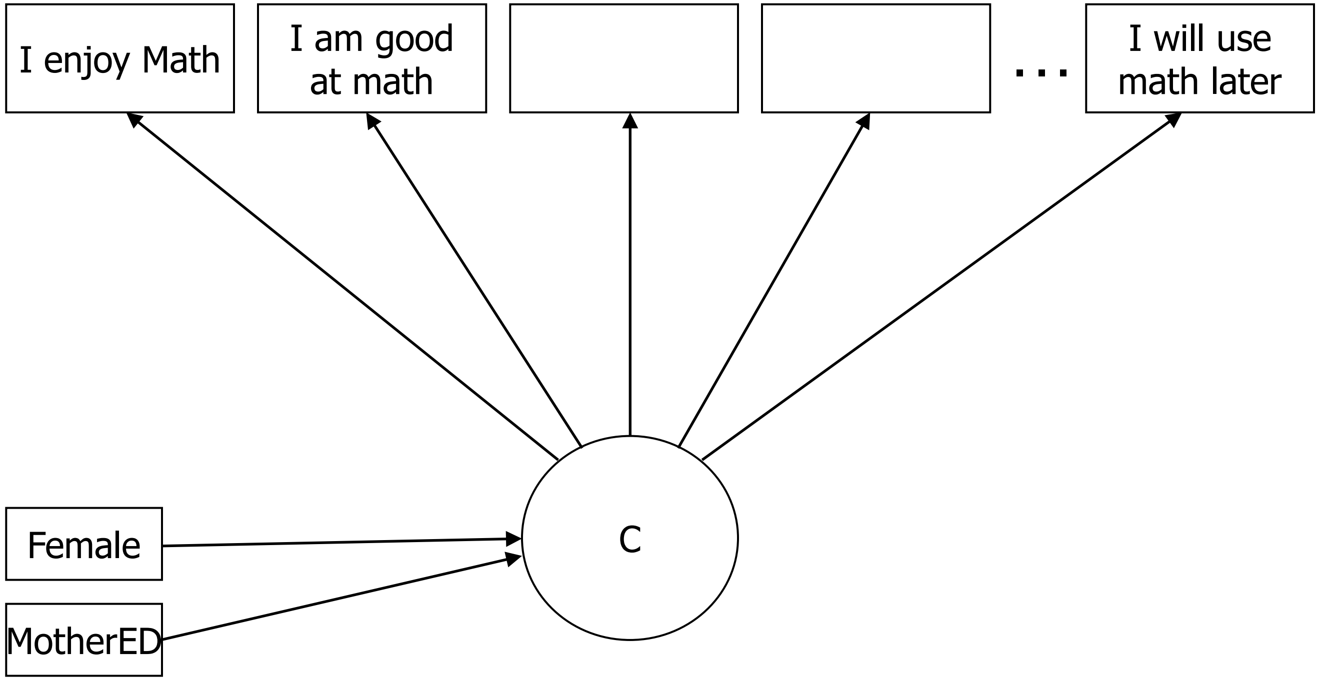
Estimated

IMMERSE Project– IES funded Training Grant (R305B220021)

ML 3-step for Latent Class Regression: Example

IMMERSE Project – IES funded Training Grant (R305B220021)

Manual ML 3-step: Latent Class regression with two covariates



IMMERSE Project– IES funded Training Grant (R305B220021)

UC SANTA BARBARA

Step 1: Fit unconditional LCA (save CPROB)

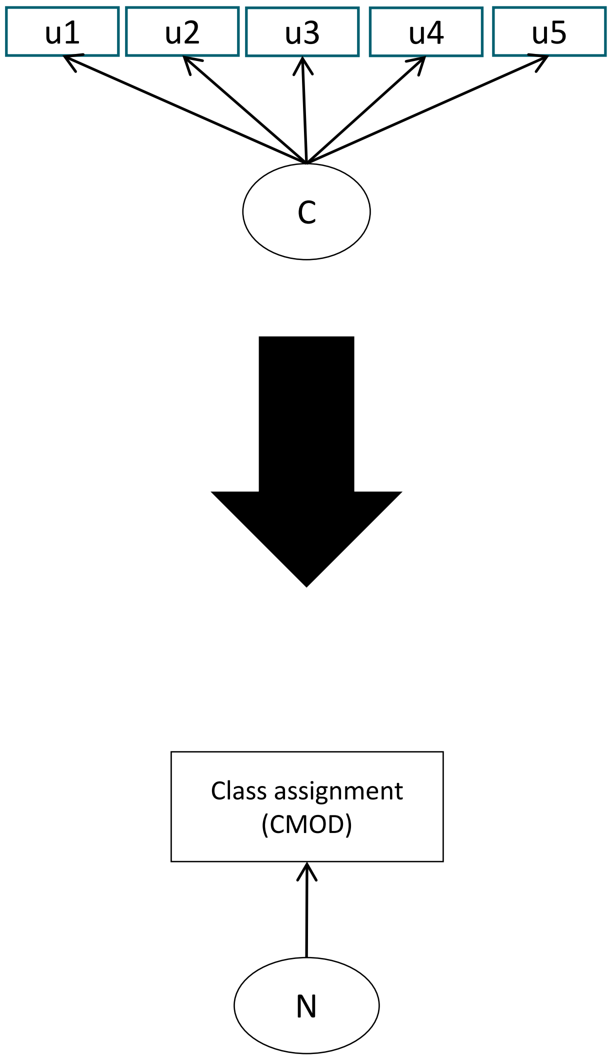
```
usevar = ca28ar ca28br ca28cr ca28er  ca28gr ca28hr ca28ir
    ca28kr ca28lr;

  CATEGORICAL = ca28ar ca28br ca28cr ca28er  ca28gr ca28hr ca28ir
    ca28kr ca28lr;

  missing=all(9999);
  idvariable = lsayid;
  classes = c(5);
  auxiliary = gender eMTHIRTN urban mothed;

Analysis: type=mixture;
  starts = 100 10;

savedata:
    file is lsay_c5_immerse.txt;
    save = cprob;
    missflag = 9999;
    format = free;
```

Add any/all covariates and distal outcomes here. Better to be generous here.

This is creating a new file with the cprobs and any variables listed in auxiliary command.

```
SAVEDATA INFORMATION
  Save file
 lsay_c5_immerse.txt

  Order of variables
    Names are
      CA28AR
      CA28BR
      CA28CR
      CA28ER
      CA28GR      Outcome variables
      CA28HR
      CA28IR
      CA28KR
      CA28LR
      GENDER
      EMTHIRTN    Auxiliary variables
      URBAN
      MOTHED
      CPROB1
      CPROB2
      CPROB3      Posterior probabilities
      CPROB4
      CPROB5
      C           Modal class assignment
      LSAYID;     ID variable (in case you want to merge)

  Save file format            Free

  Save file record length     10000
```

Step 1



Done

Steps 2 and 3

Now we're going to do this

IMMERSE Project– IES funded Training Grant (R305B220021)

Step 2a: Save Data from Step 1

Output from Step 1

```
SAVEDATA INFORMATION
  Save file
    lsay_c5_nosplit_cmod.txt

  Order of variables
    Names are
      CA28AR
      CA28BR
      CA28CR
      CA28ER
      CA28GR
      CA28HR
      CA28IR
      CA28KR
      CA28LR
      GENDER
      EMTHIRTN
      URBAN
      MOTHED
      CPROB1
      CPROB2
      CPROB3
      CPROB4
      CPROB5
      C
      LSAYID;

  Save file format          Free
```

- CA28BR–CA28LR: Outcome variables
- GENDER, EMTHIRTN, URBAN, MOTHED: Auxiliary variables
- CPROB1–CPROB5: Posterior probabilities
- C: Modal class assignment
- LSAYID: ID variable (in case you want to merge)

Read in Data in Step 3…

```
Variable:
Names are CA28AR CA28BR CA28CR
CA28ER CA28GR  CA28HR  CA28IR
CA28KR CA28LR GENDER EMTHIRTN URBAN
MOTHED CPROB1 CPROB2 CPROB3 CPROB4
CPROB5 CMOD5 LSAYID;
```

Note: *I changed the name of the variable as I read it in. I do this to not confuse the modal class C (from step 1) from the latent class variable.*

You do not have to do this.. I just did it to keep things organized.

This does NOT change data, just what we call it in our variable list….

Step 2b: Read in savedata file (Partial Input)

Logits for the Classification Probabilities for the M
by Latent Class (Row)

	1	2	3	4	5
1	12.075	7.715	8.980	8.232	0.000
2	5.794	7.126	4.438	4.295	0.000
3	1.153	0.017	3.031	0.249	0.000
4	0.176	0.188	0.504	3.197	0.000
5	-9.409	-6.599	-2.548	-2.812	0.000

```
Variable:
Names are CA28AR CA28BR CA28CR CA28ER CA28GR
CA28HR  CA28IR CA28KR CA28LR GENDER EMTHIRTN
CPROB1CPROB2 CPROB3 CPROB4 CPROB5 CMOD5 LSAYID;

   usevariables = cmod5;
   Nominal = cmod5;
   missing=all(9999);
   classes= c(5);
   idvariable = lsayid;

Analysis: type = mixture;
   starts =0;

  Define:
    female = gender EQ 1;
MODEL:
%OVERALL%
 %C#1% !pro-math w/o anxiety

    [cmod5#1@12.075];
    [cmod5#2@7.715];
    [cmod5#3@8.980];
    [cmod5#4@8.232];

 %C#2% !pro-math w/ anxiety

    [cmod5#1@5.784];
    [cmod5#2@7.126];
    [cmod5#3@4.438];
    [cmod5#4@4.295];
```


IMMERSE Project– IES funded Training Grant (R305B220021)

Step 2c: Fix error rate values based on Step 1 (Partial Input)

```
ANALYSIS: type = mixture;
   starts =0;

Define:
    female = gender EQ 1;
MODEL:

%OVERALL%

%C#1% !pro-math w/o anxiety

    [cmod5#1@12.075];
    [cmod5#2@7.715];
    [cmod5#3@8.980];
    [cmod5#4@8.232];

%C#2% !pro-math w/ anxiety

    [cmod5#1@5.784];
    [cmod5#2@7.126];
    [cmod5#3@4.438];
    [cmod5#4@4.295];

%C#3% !math lover

    [cmod5#1@1.153];
    [cmod5#2@0.017];
    [cmod5#3@3.031];
    [cmod5#4@0.249];

%C#4% !I don't like math but know it's good for me

    [cmod5#1@0.176];
    [cmod5#2@0.188];
    [cmod5#3@0.504];
    [cmod5#4@3.197];

%C#5% !anti-math w/ anxiety

    [cmod5#1@-9.409];
    [cmod5#2@-6.599];
    [cmod5#3@-2.548];
    [cmod5#4@-2.812];
```

Logits for the Classification Probabilities for the M
by Latent Class (Row)

	1	2	3	4	5
1	12.075	7.715	8.980	8.232	0.000
2	5.794	7.126	4.438	4.295	0.000
3	1.153	0.017	3.031	0.249	0.000
4	0.176	0.188	0.504	3.197	0.000
5	-9.409	-6.599	-2.548	-2.812	0.000

There is R code for Mplus Automation to do this. (Garber, 2021)

Garber, A. C. (2021, September 3). 3-Step ML Auxiliary Variable Integration Using MplusAutomation. https://doi.org/10.31234/osf.io/phtxa

UC SANTA BARBARA

Step 2d: Run model and check to see if it worked

Class counts from Step 1

```
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL

    Latent
   Classes

       1        1059.25081          0.39598
       2         331.28420          0.12384
       3         569.14744          0.21277
       4         434.78364          0.16254
       5         280.53391          0.10487


FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES

    Latent
   Classes

       1        1059.25072          0.39598
       2         331.28426          0.12384
       3         569.14756          0.21277
       4         434.78344          0.16254
       5         280.53402          0.10487


FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

    Latent
   Classes

       1              1128          0.42168
       2               290          0.10841
       3               538          0.20112
       4               437          0.16336
       5               282          0.10542
```

Class counts from Step 2d

```
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL

    Latent
   Classes

       1        1059.94299          0.39624
       2         330.67478          0.12362
       3         569.10508          0.21275
       4         434.73592          0.16252
       5         280.54123          0.10488


FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES

    Latent
   Classes

       1        1059.94369          0.39624
       2         330.67319          0.12362
       3         569.10563          0.21275
       4         434.73651          0.16252
       5         280.54099          0.10488


FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

    Latent
   Classes

       1              1128          0.42168
       2               290          0.10841
       3               538          0.20112
       4               437          0.16336
       5               282          0.10542
```


IMMERSE Project– IES funded Training Grant (R305B220021)

Checking in

- So far we have specified the manual 3 step.
- Have NOT done any auxiliary variable analysis yet.
- Now the fun begins

Step 3: Latent class regression

```
   usevariables = cmod5  mothed female;
   Nominal = cmod5;
    missing=all(9999);
    classes= c(5);

    idvariable = lsayid;

Analysis: type = mixture;
 starts =0;

Define:
  female = gender EQ 1;

Model:
  %OVERALL%
  c on female (f1-f4);
  c on mothed (m1-m4);
```

Latent class regression

```
%C#1% !pro-math w/o anxiety

  [cmod5#1@12.075];
  [cmod5#2@7.715];
  [cmod5#3@8.980];
  [cmod5#4@8.232];

  %C#2% !pro-math w/ anxiety

  [cmod5#1@5.784];
  [cmod5#2@7.126];
  [cmod5#3@4.438];
  [cmod5#4@4.295];

%C#3% !math lover

  [cmod5#1@1.153];
  [cmod5#2@0.017];
  [cmod5#3@3.031];
  [cmod5#4@0.249];

. . .
```

This is the same as we did in previous slide

IMMERSE Project– IES funded Training Grant (R305B220021)

Hypothesis Testing with LCR

- Same as what we did with multinomial regression!
- Overall test of an association between a given predictor and latent class membership requires the omnibus test.
- Is there evidence of an overall association, then we can probe for where the differences are using pairwise comparison methods.

Omnibus Tests of Association

```
Model:
  %OVERALL%
  c on female mothed;
```

Regressing C on covariates

```
Model:
  %OVERALL%
  c on female (f1-f4);
  c on mothed (m1-m4);
```

Regressing C on covariates, adding labels for the logit slopes. In this case there are five classes, thus four logit equations.

f1-f4 are the slopes for female in the logits.
m1-m4 are the slope of mothed in the logits.

IMMERSE Project– IES funded Training Grant (R305B220021)

UC SANTA BARBARA

Omnibus Tests of Association

```
Model:
  %OVERALL%
  c on female (f1-f4);
  c on mothed (m1-m4);
.
.
.
Model Test:
    0=f1-f2;
    0=f1-f3;
    0=f1-f4;

    !0=m1-m2;
    !0=m1-m3;
    !0=m1-m4;
```

Model Test will test if all statements are simultaneously true.

***Note*, you have to run this code twice. One testing f1-f4 and then again, running m1-m4**

```
Wald Test of Parameter Constraints

          Value                              29.380
          Degrees of Freedom                      3
          P-Value                            0.0000
```

What does this result mean about the relation between female and the latent class variable?

Omnibus Tests of Association

Female

```
Wald Test of Parameter Constraints

          Value                              29.380
          Degrees of Freedom                      3
          P-Value                            0.0000
```

```
Model Test:
 0=f1-f2;
 0=f1-f3;
 0=f1-f4;

 !0=m1-m2;
 !0=m1-m3;
 !0=m1-m4;
```

There is a statistically significant overall association between each of the two covariates–gender ($X2 = 29.380$, *df* = 3, *p*<.001) and mother's education ($X2 = 329.380$, *df* = 3, *p*<.001)–and the latent class variable of math deposition accounting for the other.

Mother's Education

```
Wald Test of Parameter Constraints

          Value                              29.380
          Degrees of Freedom                      3
          P-Value                            0.0000
```

```
Model Test:
  !0=f1-f2;
  !0=f1-f3;
  !0=f1-f4;

  0=m1-m2;
  0=m1-m3;
  0=m1-m4;
```

ML 3-step: LCR

Categorical Latent Variables

C#1 ON				
FEMALE	0.258	0.157	1.643	0.100
MOTHED	0.066	0.077	0.858	0.391
C#2 ON				
FEMALE	-0.279	0.208	-1.338	0.181
MOTHED	-0.023	0.102	-0.223	0.823
C#3 ON				
FEMALE	0.200	0.189	1.057	0.291
MOTHED	-0.028	0.096	-0.287	0.774
C#4 ON				
FEMALE	0.818	0.198	4.139	0.000
MOTHED	-0.161	0.100	-1.619	0.106
Intercepts				
C#1	1.040	0.213	4.882	0.000
C#2	0.323	0.273	1.183	0.237
C#3	0.687	0.262	2.621	0.009
C#4	0.352	0.275	1.279	0.201

LOGISTIC REGRESSION ODDS RATIO RESULTS

			95% C.I.	
	Estimate	S.E.	Lower 2.5%	Upper 2.5%
Categorical Latent Variables				
C#1 ON				
FEMALE	1.295	0.204	0.951	1.762
MOTHED	1.068	0.082	0.919	1.241
C#2 ON				
FEMALE	0.757	0.158	0.503	1.138
MOTHED	0.978	0.100	0.801	1.193
C#3 ON				
FEMALE	1.221	0.231	0.843	1.768
MOTHED	0.973	0.094	0.805	1.175
C#4 ON				
FEMALE	2.267	0.448	1.538	3.339
MOTHED	0.851	0.085	0.700	1.035

ML 3-step: LCR

Mplus provides the estimates changing the reference class.

What do you notice in terms of significant patterns compared to the default comparison?

ALTERNATIVE PARAMETERIZATIONS FOR THE CATEGORICAL LATENT VARIABLE REGRESSION

Parameterization using Reference Class 1

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
C#2 ON				
MOTHED	-0.089	0.087	-1.020	0.308
FEMALE	-0.537	0.179	-2.998	0.003
C#3 ON				
MOTHED	-0.093	0.071	-1.321	0.187
FEMALE	-0.059	0.140	-0.419	0.675
C#4 ON				
MOTHED	-0.227	0.073	-3.093	0.002
FEMALE	0.560	0.147	3.801	0.000
C#5 ON				
MOTHED	-0.066	0.077	-0.858	0.391
FEMALE	-0.258	0.157	-1.643	0.100
Intercepts				
C#2	-0.717	0.230	-3.111	0.002
C#3	-0.353	0.192	-1.837	0.066
C#4	-0.688	0.204	-3.380	0.001
C#5	-1.040	0.213	-4.882	0.000

Parameterization using Reference Class 2

C#1 ON				
MOTHED	0.089	0.087	1.020	0.308
FEMALE	0.537	0.179	2.998	0.003
C#3 ON				
MOTHED	-0.005	0.100	-0.049	0.961
FEMALE	0.478	0.198	2.410	0.016
C#4 ON				
MOTHED	-0.138	0.105	-1.312	0.190
FEMALE	1.097	0.209	5.259	0.000
C#5 ON				
MOTHED	0.023	0.102	0.223	0.823
FEMALE	0.279	0.208	1.338	0.181
Intercepts				
C#1	0.717	0.230	3.111	0.002
C#3	0.364	0.264	1.381	0.167
C#4	0.029	0.281	0.102	0.919
C#5	-0.323	0.273	-1.183	0.237

IMMERSE Project– IES funded Training Grant (R305B220021)

ML 3-step: LCR

Categorical Latent Variables

C#1 ON				
FEMALE	0.258	0.157	1.643	0.100
MOTHED	0.066	0.077	0.858	0.391
C#2 ON				
FEMALE	-0.279	0.208	-1.338	0.181
MOTHED	-0.023	0.102	-0.223	0.823
C#3 ON				
FEMALE	0.200	0.189	1.057	0.291
MOTHED	-0.028	0.096	-0.287	0.774
C#4 ON				
FEMALE	0.818	0.198	4.139	0.000
MOTHED	-0.161	0.100	-1.619	0.106
Intercepts				
C#1	1.040	0.213	4.882	0.000
C#2	0.323	0.273	1.183	0.237
C#3	0.687	0.262	2.621	0.009
C#4	0.352	0.275	1.279	0.201

LOGISTIC REGRESSION ODDS RATIO RESULTS

	Estimate	S.E.	95% C.I. Lower 2.5%	Upper 2.5%
Categorical Latent Variables				
C#1 ON				
FEMALE	1.295	0.204	0.951	1.762
MOTHED	1.068	0.082	0.919	1.241
C#2 ON				
FEMALE	0.757	0.158	0.503	1.138
MOTHED	0.978	0.100	0.801	1.193
C#3 ON				
FEMALE	1.221	0.231	0.843	1.768
MOTHED	0.973	0.094	0.805	1.175
C#4 ON				
				3.339
				1.035

Interpretation:
Significant logit: difference in the log odds for one unit increase in "female" variable. Females, compared to males, have a significantly higher log odds (logit= .818, *p*<.01) of being in C4 relative to C5. The odds of being in C4 relative to C5 for females, compared to males, is 2.267 (95% CI 1.53, 2.39)

```
C#1 !pro-math w/o anxiety
C#2 !pro-math w/ anxiety
C#3 !math lover
C#4 !I don't like math but know it's good for me
C#5 !anti-math w/ anxiety
```

Sample table and write-up

Table 3
Log odds coefficients and odds ratio for the five-class model with age, kindergarten CELDT scores, K special education status, gender, and preschool experience as predictors using the balanced, high class as the comparison group.

Readiness class	Effect	Logit	SE	t	Odds ratio
Moderate SE, Moderate Cog	Age	0.37	0.49	0.77	1.45
	K CELDT	−0.05	0.09	−0.62	0.95
	K Special Ed	−0.33	0.95	−0.35	0.72
	Female	−0.96**	0.36	−2.66	0.38
	Preschool	0.23	0.51	0.45	1.25
Moderate SE, Low Cog	Age	−0.52	0.51	−1.03	0.59
	K CELDT	−0.67**	0.10	−6.67	0.51
	K Special Ed	−1.17	3.39	−0.35	0.31
	Female	0.46	0.46	0.99	1.58
	Preschool	−1.73**	0.61	−2.86	0.18
Low SE, Low Cog	Age	−0.20	0.41	−0.47	0.82
	K CELDT	−0.34*	0.17	−1.96	0.72
	K Special Ed	0.93	0.76	1.22	2.53
	Female	−1.01*	0.43	−2.32	0.37
	Preschool	−1.68**	0.55	−3.05	0.19
Extremely Low	Age	−1.12**	0.45	−2.46	0.33
	K CELDT	−0.51**	0.09	−5.76	0.60
	K Special Ed	1.46	0.97	1.51	4.32
	Female	−0.34	0.37	−0.93	0.71
	Preschool	−3.98*	0.56	−7.06	0.02

Note: SE = social-emotional, Cog = cognitive.
* $p < 0.05$.
** $p < 0.01$.

5.2. Examining predictors and distal outcomes

Table 2 presents descriptive information of the predictors and distal outcomes used in the analysis. The latent class variable was regressed on to all of the predictors included in the model. Since the latent class variable is a categorical latent variable, the regression of this variable on the predictors was a multinomial logistic regression, and instead of interpreting regression coefficients, we interpreted logits. We chose the students in the *Balanced High* class to be the comparison group, and compared the other four classes to this group on each of the predictors. Table 3 presents the logit parameters, their standard errors, the corresponding *t*-value, and the odds ratio for each comparison.

Comparing students in the *Mod SE, Mod Cog* class to the *Balanced High* class, there was no significant difference in age, English proficiency (−0.50, $p > 0.05$, $OR = 0.95$), kindergarten special education placement (−0.33, $p > 0.05$, $OR = 0.72$), or preschool experience. That is, students in these two classes have similar ages and English proficiency, and had the same proportion of students with special education placements and preschool experience. There was a significant gender effect (−0.96, $p < 0.05$, $OR = 0.38$) — girls were significantly less likely to be in the *Mod SE, Mod Cog* class compared to the *Balanced High* class.

Children in the *Mod SE, Low Cog* class were similar to those in the *Balanced High* class with respect to age (−0.52, $p > 0.05$, $OR = 0.59$) and gender (0.46, $p > 0.05$, $OR = 1.58$). Though not statistically significant, there was a notably small odds ratio for special education (−1.17, $p > 0.05$, $OR = 0.31$). There were significant differences with respect to English proficiency (−0.67, $p < 0.05$, $OR = 0.51$) and preschool experience (−1.73, $p < 0.05$, $OR = 0.19$). These results indicated that children in the *Mod SE, Low Cog* class had significantly lower English proficiency scores and had less children exposed to preschool, and had lower odds of placement in special education compared to those in the *Balanced High* class.

Comparing children in the *Low SE, Low Cog* class to the *Balanced High* class, there was no difference with respect to age (−0.20, $p > 0.05$, $OR = 0.59$). There were significant differences with respect to English proficiency (−0.34, $p < 0.05$, $OR = 0.72$), gender (−1.01, $p < 0.05$, $OR = 0.37$), and preschool exposure (−1.68, $p < 0.05$, $OR = 0.19$). There was a notably large odds ratio for special education, though not statistically significant (0.93, $p > 0.05$, $OR = 2.53$). That is, children in the *Low SE, Low Cog* class had significantly lower English proficiency scores, were less likely to be females, had higher

How could this table be simplified?

Is there anything missing from the table?

Quirk, M., Nylund-Gibson, K., & Furlong, M. (2013). Exploring patterns of Latino/a children's school readiness at kindergarten entry and their relations with Grade 2 achievement. *Early Childhood Research Quarterly*, *28*(2), 437-449.

Sample table

Table 4

Logits and Odds Ratios (OR) of the Predictors of Class Membership by Reference Class

Class membership	Effect	Reference class: Minimal		Peer Victims		Poly Victims		Poly (sexual)	
		Logit	*OR*	Logit	*OR*	Logit	*OR*	Logit	*OR*
Minimal	Female			.06	1.07	−.84**	.43	−1.53*	.22
	LGBTQ			−.77	.47	−1.22*	.29	−.71	.49
	Latino			1.01**	2.75	1.11**	3.04	−.34	.71
	Other			.18	1.20	.35	1.42	−.35	.70
	Fall depression			.02	1.02	−.12**	.89	−.12	.89
	Fall anxiety			−.09	.92	−.08	.92	−.08	.92
Peer Victims	Female	−.06	.94			−.91**	.40	−1.59*	.20
	LGBTQ	.77	2.15			−.46	.63	.05	1.05
	Latino	−1.01**	.36			.10	1.11	−1.35*	.26
	Other	−.18	.84			.17	1.19	−.53	.59
	Fall depression	−.02	.98			−.14**	.87	−.14*	.87
	Fall anxiety	.09	1.09			.01	1.01	.01	1.01
Poly Victims	Female	.84**	2.32	.91**	2.48			−.68	.51
	LGBTQ	1.22*	3.39	.46	1.58			.51	1.67
	Latino	−1.11**	.33	−.10	.90			−1.45*	.23
	Other	−.35	.00	−.17	.84			−.70	.50
	Fall depression	.12**	1.13	.14**	1.15			.01	1.01
	Fall anxiety	.08	1.08	−.01	.99			.01	1.01
Poly (sexual)	Female	1.53*	4.60	1.59*	4.89	.68	1.98		
	LGBTQ	.71	2.04	−.05	.95	−.51	.60		
	Latino	.34	1.40	1.35*	3.85	1.45*	4.26		
	Other	.35	1.41	.53	1.69	.70	2.01		
	Fall depression	.12	1.12	.14*	1.15	.00	1.00		
	Fall anxiety	.08	1.08	−.01	.99	.00	1.00		

Note. LGBTQ = lesbian, gay, bisexual, transgender, queer.
* $p < .05$ ** $p < .01$ *** $p < .001$

How could this table be simplified?

Is there anything missing from the table?

Holt, M. K., Felix, E., Grimm, R., Nylund-Gibson, K., Green, J. G., Poteat, V. P., & Zhang, C. (2017). A latent class analysis of past victimization exposures as predictors of college mental health. *Psychology of violence*, *7*(4), 521.

Automatic ML 3-Step

Just FYI…

Automatic ML 3-step (covariate)

Embedded in Mplus and limited to only covariates or only distal outcomes.

R3step – latent class regression using ML 3-step

```
usevariables = ca28ar ca28br ca28cr ca28er  ca28gr ca28hr ca28ir
   ca28kr ca28lr;
CATEGORICAL = ca28ar ca28br ca28cr ca28er  ca28gr ca28hr ca28ir
   ca28kr ca28lr;

   Auxiliary  = gender (r3step);
   idvariable = lsayid;
   missing=all(9999);

   classes=c(5);

 Analysis:
   type= mixture;
   starts=100 20;
```

← Gender is the covariate

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
C#1 ON				
GENDER	-0.297	0.156	-1.904	0.057
C#2 ON				
GENDER	0.284	0.208	1.367	0.172
C#3 ON				
GENDER	-0.245	0.188	-1.306	0.192
C#4 ON				
GENDER	-0.864	0.195	-4.430	0.000
Intercepts				
C#1	1.787	0.258	6.934	0.000
C#2	-0.292	0.355	-0.823	0.411
C#3	1.088	0.309	3.522	0.000
C#4	1.714	0.306	5.611	0.000

IMMERSE Project– IES funded Training Grant (R305B220021)

Lab?