Advanced Psychometrics Methods Workshop

Longitudinal Data Analysis in 1 hour (optional)

Rich Jones
rnjones@brown.edu @rnjma

Granlibakken Conference Center
Lake Tahoe, CA August 21, 2018

Objective

Resources to learn more about LDA and LDA using SEM methods

Introduce some of the concepts and terminology relevant to longitudinal data analysis (LDA)

with special emphasis on applications in cognitive aging research using structural equation modeling (SEM)

Acknowledgements

- Funded in part by Grant R13AG030995 from the National Institute on Aging
- The views expressed in written conference materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention by trade names, commercial practices, or organizations imply endorsement by the U.S. Government

Session Overview

1. Other resources
2. General Modeling Framework
3. Example
4. Questions and discussion

Other Resources

- **What is longitudinal data analysis?**
  - Singer JD Willett JB. Applied longitudinal data analysis: Modeling change and event occurrence. 2003, New York: OUP.
  - Data and worked examples at UCLA Institute for Digital Research and Education (IDRE) web site https://stats.idre.ucla.edu/other/examples/alda/

Other Resources

- **How do I do LDA, especially these SEM Approaches?**
  - Newsom J, Jones RN, Hofer S (Eds). Longitudinal Data Analysis: A Practical Guide for Researchers in Aging, Health and Social Sciences. 2011. New York: Routledge.
  - Duncan TE, Duncan SC, Strycker LA. An introduction to latent variable growth curve modeling: concepts, issues and applications. Second ed. 2006, Mahwah, NJ: LEA, Inc.
  - Mirman, D. (2013). Growth curve analysis and visualization using R. CRC Press.
  - Beaujean, A. A. (2014). Latent variable modeling using R: A step-by-step guide. Routledge.

Other Resources

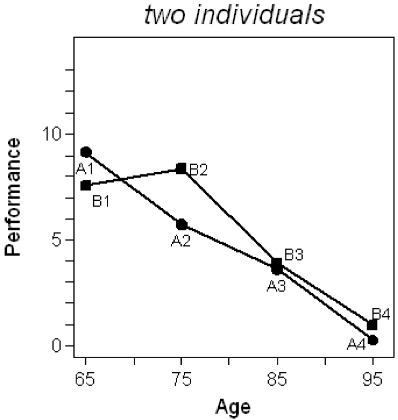
- **Tell me more about the math behind latent curve methods**
  - Bollen KA, Curran PJ. Latent curve models: a structural equation perspective. 2006, Hoboken, NJ.: Wiley

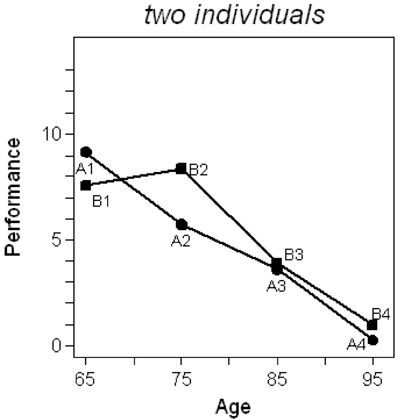
Other Resources

- **Latent Variable Methods Workshop** in Providence
  - Measurement models - even years (e.g., 2020)
  - Longitudinal data analysis - odd years (e.g., 2019)
  - For information
  - For LDA slides, examples (code and data)
- Other courses listed on the Mplus web site.

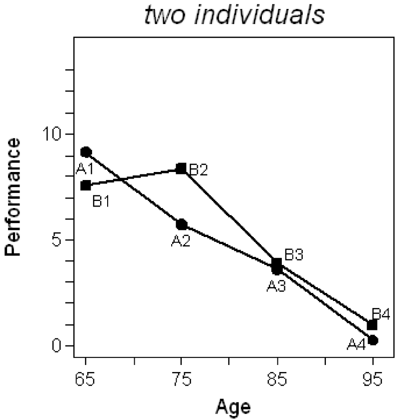
What is Longitudinal Data Analysis (LDA)?

- Analysis of data where observations are repeated or replicated
- More interesting if more than 2 observations (generally)
- Outcomes can be
  - absorbing events
    - e.g., death, conversion to dementia
  - discrete but non-absorbing
    - e.g., conversion to mild cognitive impairment
  - quantitatively distributed
    - e.g., neuropsychological test performance
  - quantities not directly observed but measured indirectly
    - e.g., latent variables

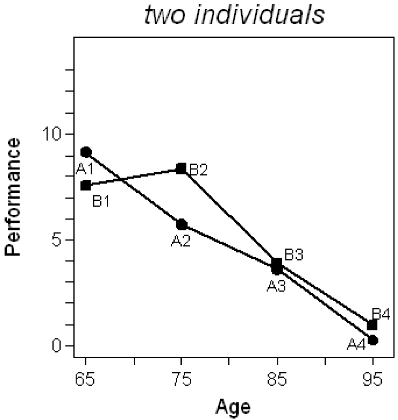

two individuals
Performance
Age
0
5
10
65
75
85
95
A1
A2
A3
A4
B1
B2
B3
B4



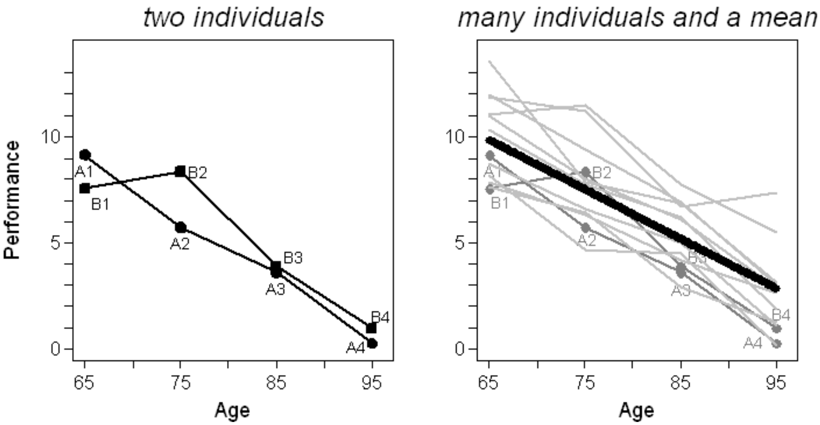
Performance =
$b_0 + b_1\text{Age} + e$

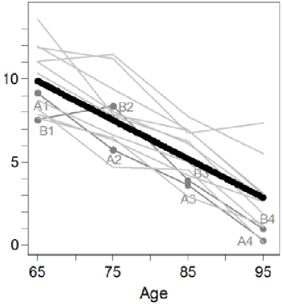

two individuals
Performance
10
5
0
A1
B1
B2
A2
B3
A3
B4
A4
65
75
85
95
Age

$$y = b_0 + b_1 x + e$$

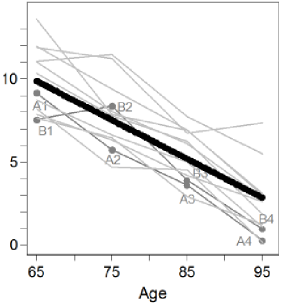


$$y_{ij} = b_{0i} + b_{1i}x_i + e_{ij}$$

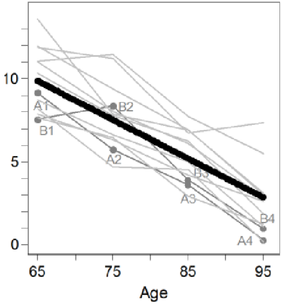

two individuals
many individuals and a mean
Performance
Age
0
5
10
65
75
85
95
A1
A2
A3
A4
B1
B2
B3
B4



$$y_{ij} = b_{0i} + b_{1i}x_j + e_{ij}$$
$$b_{0i} = a_0 + \zeta_{0i}$$
$$b_{1i} = a_1 + \zeta_{1i}$$
$$i \in [1, N]$$
$$j \in [1, T]$$
$$e \sim N(0, \theta), COV(y, \theta) = 0$$
$$\zeta \sim N(0, \psi), COV(y, \zeta) = 0$$



$$
\begin{aligned}
y_{ij} &= \eta_{1i} \times 1 \quad && + \eta_{2i}\lambda_j + \epsilon_{ij} \\
y_{ij} &= \eta_{1i} && + \eta_{2i}\lambda_j + \epsilon_{ij} \\
\eta_{1i} &= \alpha_1 + \zeta_1 \\
\eta_{2i} &= \alpha_2 + \zeta_2 \\
i &\in [1, N] \\
j &\in [1, T] \\
\epsilon &\sim N(0, \theta), \quad \zeta \sim N(0, \psi)
\end{aligned}
$$



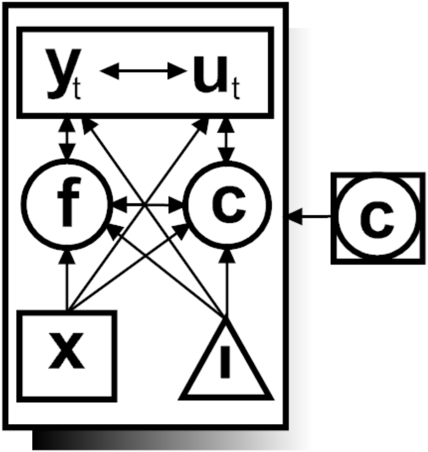
$$y_{ij} = \mathrm{i}_i + \mathrm{s}_i \lambda_j + e_{ij}$$
$$\mathrm{i}_i = \alpha(i) + \zeta(i)_i$$
$$\mathrm{s}_i = \alpha(s) + \zeta(s)_i$$
$$i \in [1, N]$$
$$j \in [1, T]$$
$$e \sim N(0, \theta), \zeta \sim N(0, \psi)$$

Advantages of LDA in SEM

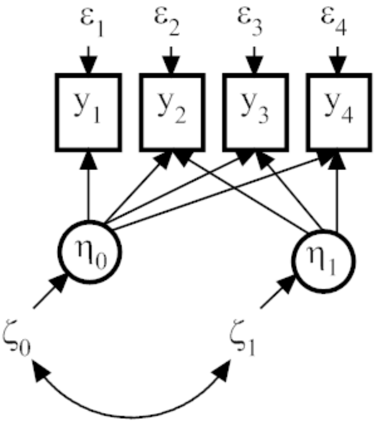
- **None**. There aren't any advantages if the analysis question is relatively straight-forward.
  - If interested in change over time, including group differences in change over time, conventional random effects or mixed effect modeling is a better choice than SEM-based approaches
  - Use trandtional random effects, estimating equations, repeated measures ANOVA if readers (and article reviewers) are more familiar with those methods

Advantages of LDA in SEM

- If you have challenging design issues, or
- Complexity that is substantively important
  - Joint models
  - Simultaneous processes
  - Sequential processes
  - Missing data modeling
  - Sub-populations (observed or latent)
  - Multi-level models
  - Weights in random effects models
  - Measurement model embedded in longitudinal model
- These, and other, model extensions can be combined

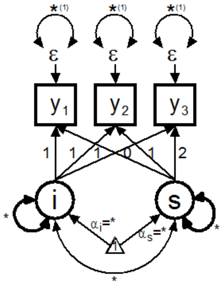

y_t
u_t
f
c
c
x
I

Simple Unconditional LGCM



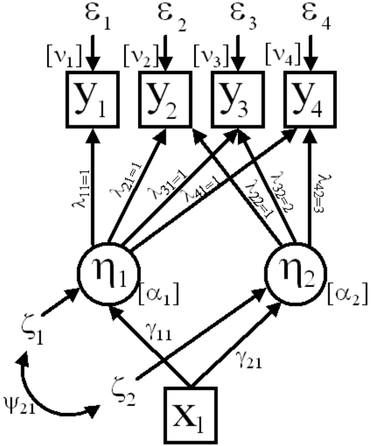
From Singer and Willet (2004) and UCLA/ATS

Model B (Table 4.1)



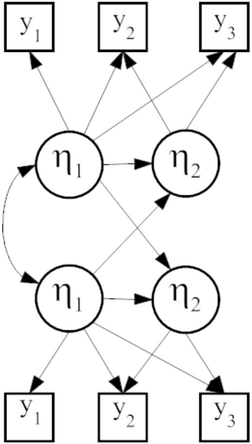

```
* SAS Example from http://www.ats.ucla.edu/stat/examples/alda.htm
*---------------------------------------------------------------------------------
* SAS
proc mixed data="c:\alda\alcohol1_pp" method=ml noclprint noinfo covtest;
  title2 "Model B";
  class id;
  model alcuse = age_14/solution notest;
  random intercept age_14/type=un sub=id;

*---------------------------------------------------------------------------------

* Mplus (short hand)
DATA:     FILE = C:\work\Shows\SHORTC~1\2009\data\swch4.dat ;
VARIABLE: NAMES = alcuse1 alcuse2 alcuse3 ;
ANALYSIS: ESTIMATOR = mlr ;
MODEL:    i s | alcuse1@0 alcuse2@1 alcuse3@2 ;
            alcuse1-alcuse3 (1) ;

*---------------------------------------------------------------------------------

* Mplus (long hand)
DATA:     FILE = C:\work\Shows\SHORTC~1\2009\data\swch4.dat ;
VARIABLE: NAMES = alcuse1 alcuse2 alcuse3 ;
ANALYSIS: ESTIMATOR = mlr ;
MODEL:    i by alcuse1-alcuse3@1 ;
          s by alcuse1@0 alcuse2@1 alcuse3@2 ;
          [alcuse1-alcuse3@0] ;
          [i* s*] ;
          alcuse1-alcuse3 (1) ;
```

Conditional LGCM



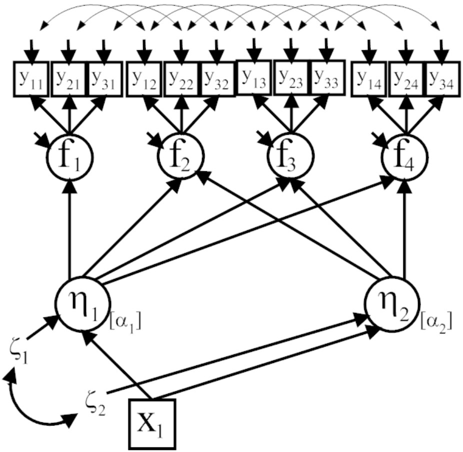
Parallel Process LGCM

Two Process Changing at the Same Time: Examine Covariation



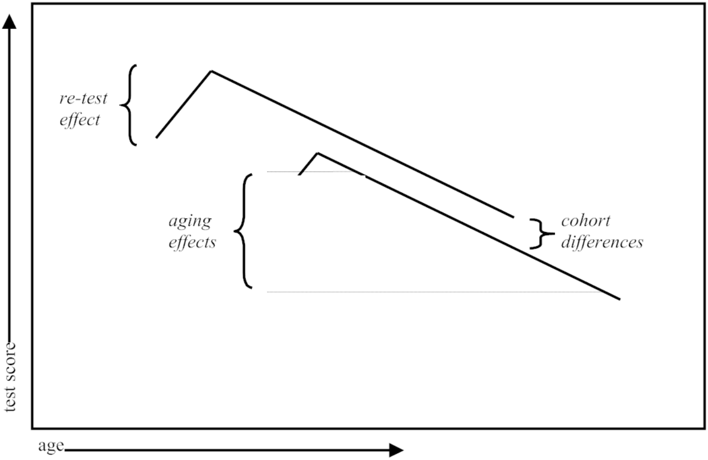
Multiple Indicator LGCM

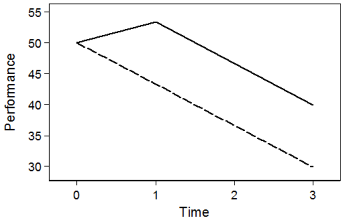
Analyze change in a latent variable by explicitly modeling it's measurement at multiple occasions (allow for DIF, missing items, other noninvariance issues)



Retest Effects

A real problem with repeat neuropsychological test adminstration.



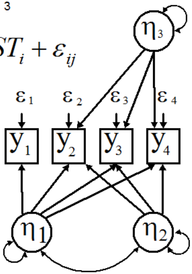


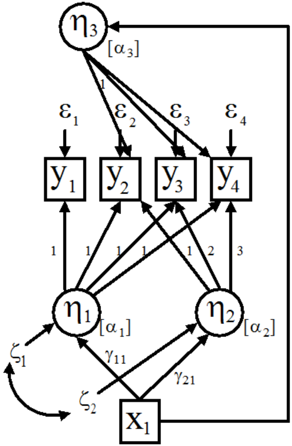
$$y_{ij} = \eta_{1i} + \eta_{2i} \times TIME_i + \eta_{3i} \times RETEST_i + \varepsilon_{ij}$$

$$\eta_{1i} = \alpha_1 + \zeta_{1i}$$

$$\eta_{2i} = \alpha_2 + \zeta_{2i}$$

$$\eta_{3i} = \alpha_3 + \zeta_{3i}$$

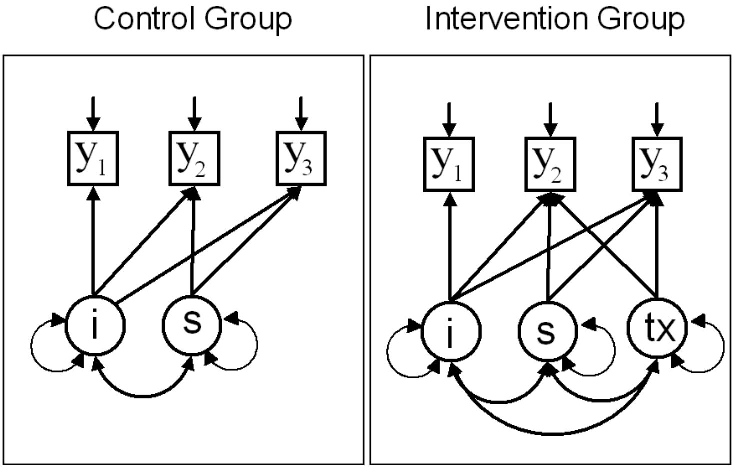





```
TITLE:     Latent Growth Curve
           with retest effect
DATA:      FILE = blah.dat ;
VARIABLE: NAMES= y1-y4 x1 ;
MODEL:
  i s | y1@0 y2@0 y3@0 y4@0 ;
  i s on x1 ;
  r by y2-y4@1 ;
  [r*] ;
  ! test what happens
  ! relaxing the constraints
  ! below...
  r@0 ;
  r with i @0 ;
  r with s @0 ;
```

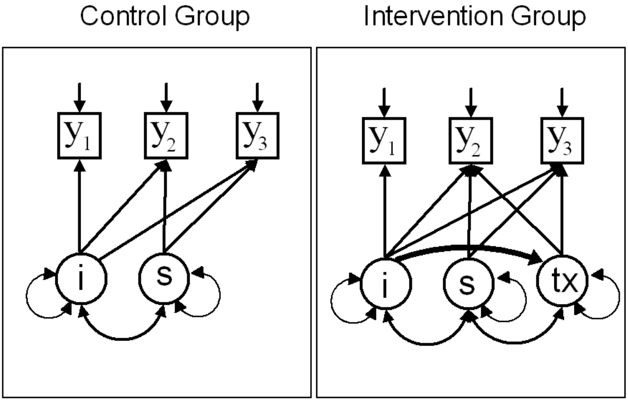
Analysis of Randomized Studies

Or natural experiments: Treatment effect as a latent variable (with a variance



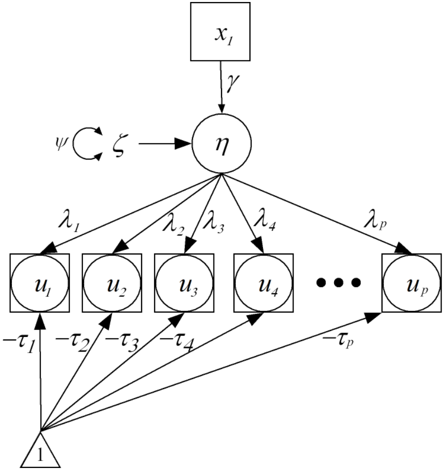
Analysis of Randomized Studies

Treatment effect dependent on baseline



Baseline-dependent treatment effect

Discrete time survival analysis



$$u_j = \begin{cases} 0 & \text{if no event at time } j \text{ \& } u_{j-1} = 0 \\ 1 & \text{if event at time } j \\ . & \text{if } u_{j-1} = 1 | u_{j-1} = . | \text{ censored at time } j \end{cases}$$

$$\boldsymbol{\tau} = \begin{pmatrix} * & * & * & * & * \end{pmatrix}$$
$$\boldsymbol{\Lambda}' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$
$$\boldsymbol{\Psi} = 0$$
$$\boldsymbol{\Gamma} = (*)$$

$$\hat{h}(j) = \frac{1}{1+e^{-\tau_j+\gamma}}$$

Joint continuous time survival and growth curve model

Mcardle et al., (2005) J. Geriatr. Psychiatry Neurol. 18(4):234

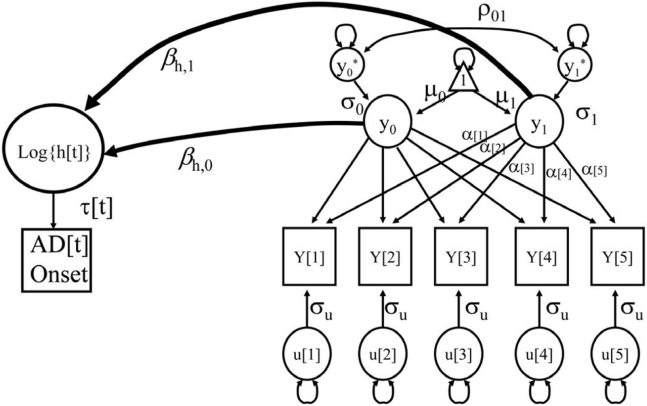
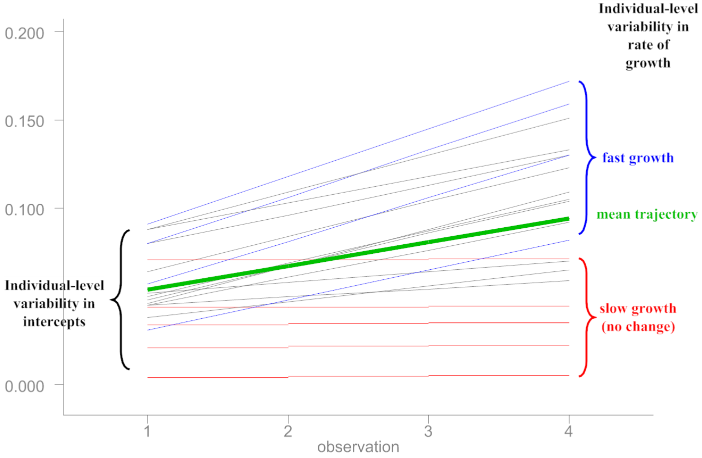
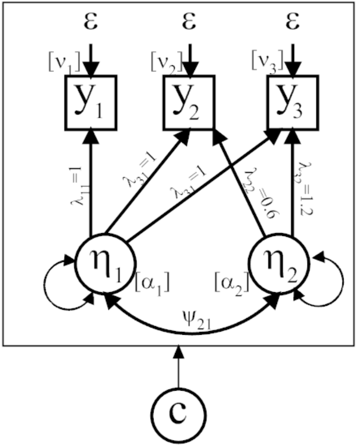
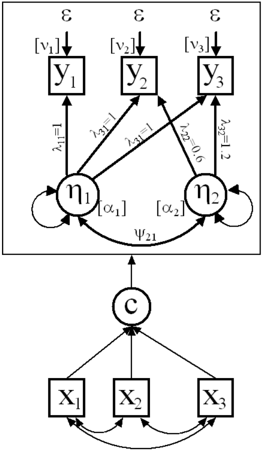


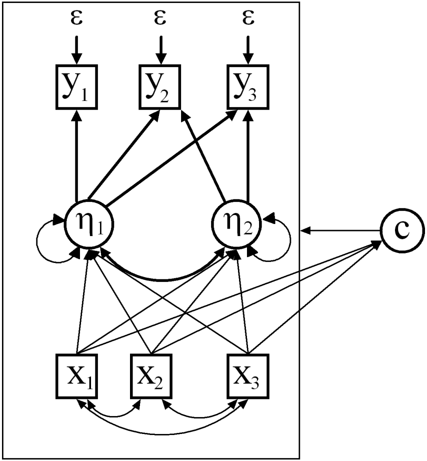
Figure 1. A path diagram of the shared-parameter concept applied to a latent growth-survival model. *Y*[t] = words recalled; *AD*[t] = Alzheimer's disease diagnosis at age *t*.

Growth Mixture Modeling




ε
ε
ε
[ν1]
[ν2]
[ν3]
y1
y2
y3
λ11=1
λ31=1
λ31=1
λ22=0.6
λ32=1.2
η1
η2
[α1]
[α2]
ψ21
c


ε
ε
ε
[ν1]
[ν2]
[ν3]
y1
y2
y3
λ11=1
λ31=1
λ31=1
λ22=0.6
λ32=1.2
η1
η2
[α1]
[α2]
ψ21
c
x1
x2
x3


ε
ε
ε
y_1
y_2
y_3
η_1
η_2
c
x_1
x_2
x_3

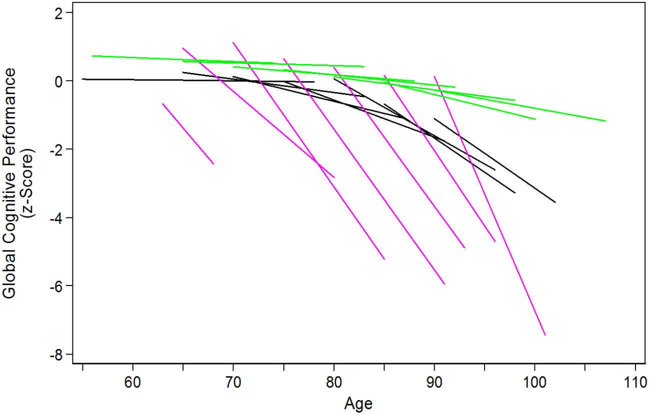
Growth Mixture Modeling

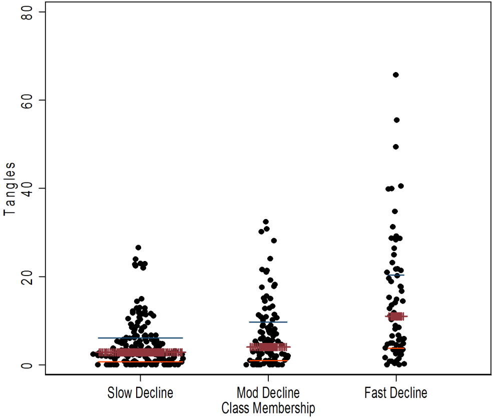
Identify population sub-samples with different growth trajectories

Age and Ageing 2011; **40:** 684–689
doi: 10.1093/ageing/afr101
Published electronically 2 September 2011

Cognitive decline in the elderly: an analysis of population heterogeneity

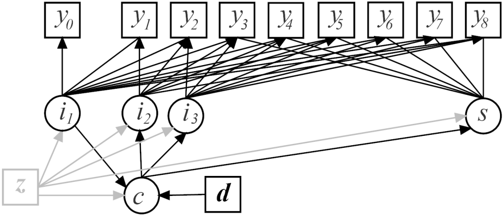
Kathleen M. Hayden[1], Bruce R. Reed[2,3], Jennifer J. Manly[4], Douglas Tommet[5], Robert H. Pietrzak[6], Gordon J. Chelune[7], Frances M. Yang[5], Andrew J. Revell[8], David A. Bennett[9], Richard N. Jones[5]




Tangles
0
20
40
60
80
Slow Decline
Mod Decline
Fast Decline
Class Membership

Growth (Knownclass) Mixture Modeling

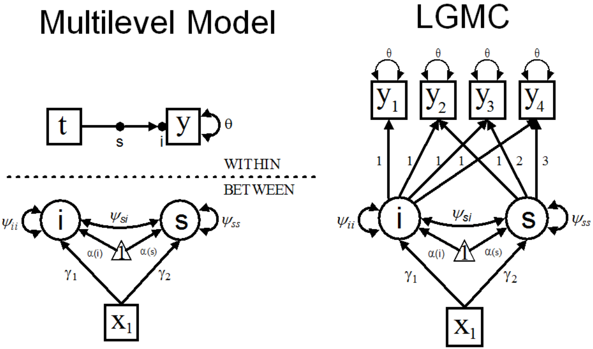
A post-baseline categorical mediator of a longitudinal trajectory



y(0)-y(8) = repeated cognitive performance assessment (0 = baseline, 8 = 36M)
i(1) = pre-operative baseline
i(2) = immediate post-operative decline (punctuation)
i(3) = immediate post-operative recovery
s = long-term slope (months 2-36)
z = background vulnerability factors, confounders
d = delirium (yes/no)
c = latent trajectory class
Note: i1, i2, i3, s all mutually correlated (not shown)

Multilevel modeling approach vs LGCM

In cases of very high intensity longitudinal data (e.g., more than 10 observations) or when there are a high number of outcomes changing simultaneously (e.g., more thatn 2 or 3), a multilevel approach can be more efficient (programmatically).

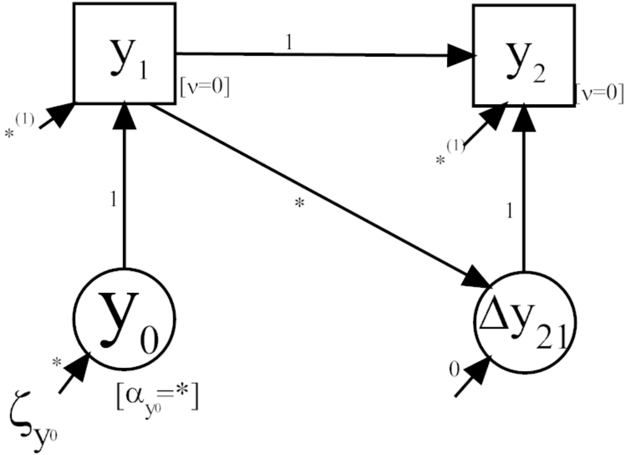


$$\mathtt{y}_{it} = \mathtt{i}_i + \mathtt{s}_i t_i + \epsilon_i \qquad \mathtt{y}_{it} = \mathtt{i}_i + \mathtt{s}_i t + \epsilon_i$$

$$\mathtt{i}_i = \alpha(\mathtt{i}) + x_{1i}\gamma_1 + \zeta(\mathtt{i})_i \qquad \mathtt{i}_i = \alpha(\mathtt{i}) + x_{1i}\gamma_1 + \zeta(\mathtt{i})_i$$

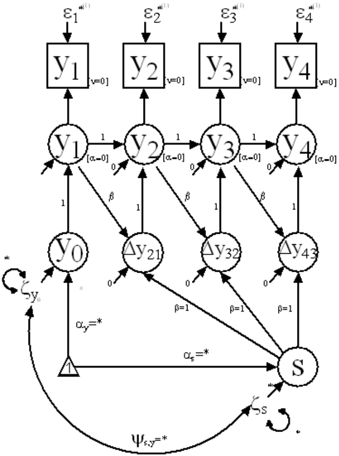
$$\mathtt{s}_i = \alpha(\mathtt{s}) + x_{1i}\gamma_2 + \zeta(\mathtt{s})_i \qquad \mathtt{s}_i = \alpha(\mathtt{s}) + x_{1i}\gamma_2 + \zeta(\mathtt{s})_i$$

Latent Difference Score Models A different approach to LDA with SEM: the change scre as a latent variable



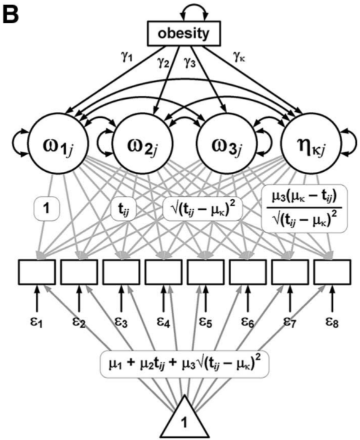
Dual Change Score Model

Change expressed with two parameters: autoregressive change score (Δ_y), and a systematic part (s). Flexible curve shapes can be estimated.



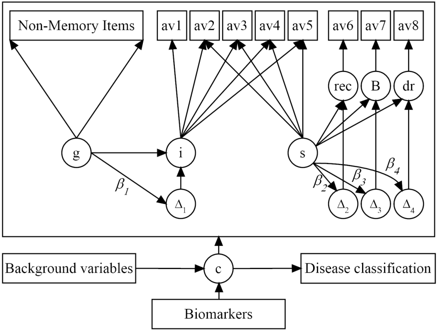
Change point models using LGCM

From: Preacher, K. J., & Hancock, G. R. (2015). Meaningful aspects of change as novel random coefficients: A general method for reparameterizing longitudinal models. Psychological methods, 20(1), 84.



A Complex Model with LGCM, LDS, Latent Classes

This is an attempt to build a better way to use neuropsychological performance data, along with biomarkers (e.g., A/T/N) and relevant background variables to classify persons into diagnostic categories. The model gives priority to modeling learning and memory (e.g., AVLT, `av1-av5`), discrepancies in word list learning and word list recognition and delayed recall trials (`av6`, `av8` via Δ_2, Δ_4) relative to other cognitive domains, and incorporating background risk factors and biomarker information.



Multilevel modeling approach + CFA for Reserve

This is a multilevel regression model. It includes a within persons model of performance on pathology. The slope from this model, a a random effect, is an indicator in a between persons model of a latent variable called *reserve* that captures covariation among random effects estimated across regressions of multiple cognitive domains on multiple pathology markers. This analytic model is a direct representation of the theoretical concept of *reserve* in a multilevel measurement model.