

2013年 10月 12日 (土) 9:30~11:30  
日本パーソナリティ心理学会第22回大会 @江戸川大学  
広報委員会企画シンポジウム

パーソナリティ心理学における統計分析の動向

# 発達や変化の軌跡をモデリングする

高橋 雄介

京都大学 大学院教育学研究科  
デザイン学大学院連携プログラム

E-mail: [takahashi.yusuke.3n@kyoto-u.ac.jp](mailto:takahashi.yusuke.3n@kyoto-u.ac.jp)





# 本報告の目的

- 縦断的に得られたデータを用いて、その特徴を十分に活かすことのできる分析を行い、横断データからは言えないことを言いたい
- 縦断研究によって得られたデータから、有用な情報を最大限引き出すためには、それらに適した統計手法を適用することが必要不可欠
- そのための統計手法として、混合軌跡モデリングを紹介する



# 縦断研究の挑戦

- 縦断研究・縦断調査・縦断データで見たいことは何か？
- 発達や変化





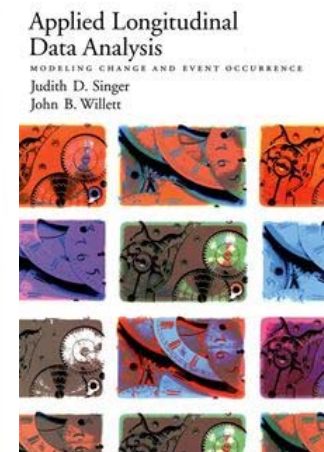
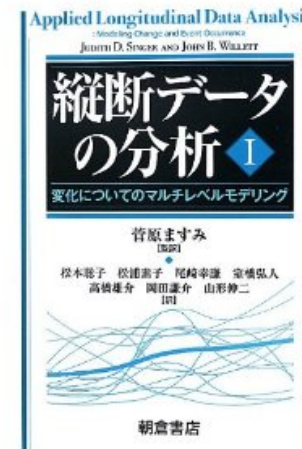
# 縦断データ・反復測定のおすすめ

- 論文の限界点に、「今回のデータは横断データなので…」、「因果関係に踏み込んだ言及を行うには縦断データが必要」と書くだけでは、済まされなくなりつつある
- 「2時点しかない」縦断データでも、横断データの何倍もモノを言える(はず)
- 実際に縦断調査をデザインして、縦断データを取得し、それらに適した統計分析にかけることが必要



# 縦断データを分析する

- 現在最も流行っていると思われる分析は？
- マルチレベルモデリング
- (階層線形モデル (HLM), 潜在成長 (曲線) モデル, 混合効果モデル)





# マルチレベル分析

- データの構造に階層性がある場合（e.g., 学校ークラスー生徒個人），まとめて分析するよりかはこの情報を分析に活かすべき，という発想
- 縦断データに対してこのモデルを適用する場合は，「クラスの中に生徒が複数いる」という階層性ではなく，「個人が複数回繰り返し測定される」“個人間差ー個人内差”という階層性を仮定する
- （今日のトークではHLMについてはこれ以上は扱いません）

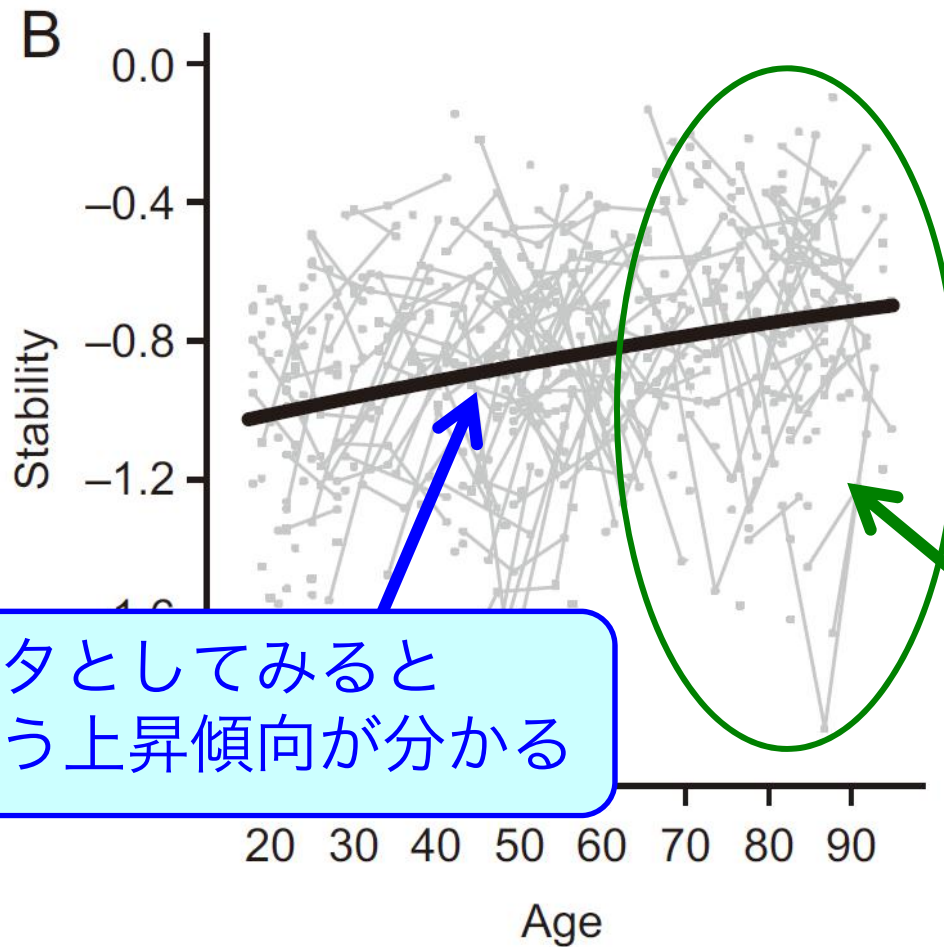


## なぜ今回はHLMは扱わないのか

- HLMにも，少なからず欠点はある
- たしかに個人差は考慮するが，背後にいくつかのグループは仮定しない
- 標本内のすべての個人が平均的な発達動態を示すとは限らない



# なぜ今回はHLMは扱わないのか



横断データとしてみると  
加齢に伴う上昇傾向が分かる

縦断データとして  
見ると経年変化に  
大きなばらつきが  
あることも分かる

(Carstensen et al., 2011, Fig1.)





# なぜ今回はHLMは扱わないのか

- 標本内にいくつかの群が存在する？
- それぞれの群が特徴をもった発達軌跡を描いている？
- 「平均グループ」は本当に存在するのか？
- ひとつの解決方法は、多母集団にすること
- 多母集団の潜在成長曲線モデル = 潜在混合モデリング (いわゆるgrowth mixture)



## 今回はなぜHLMは扱わないのか

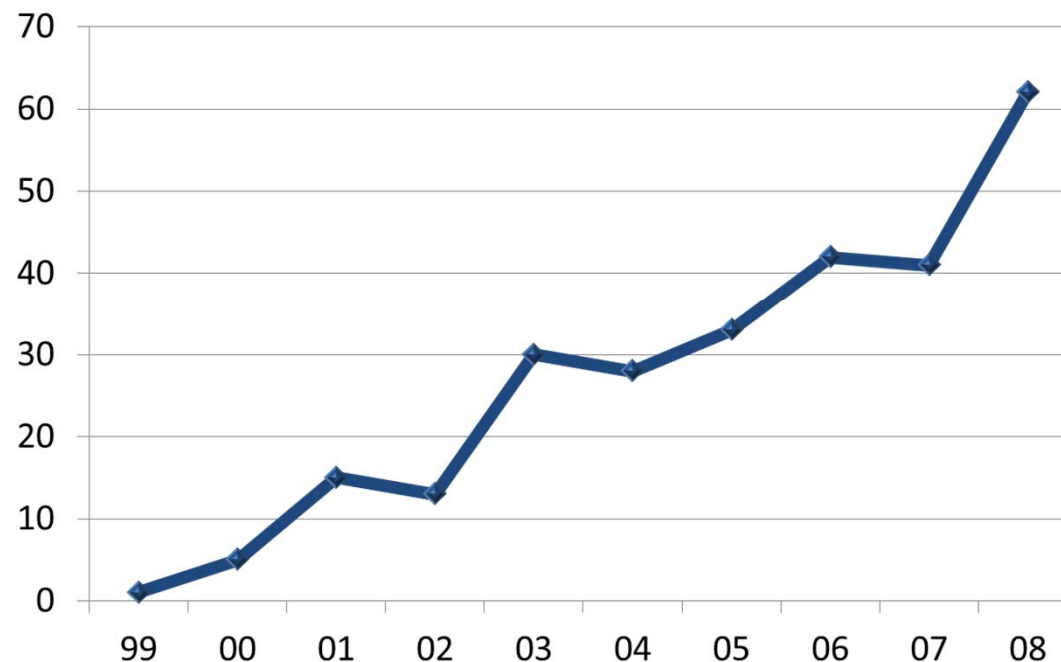
- 群レベルと個人レベルの両方のパラメータ推定が必要となるため、モデルが複雑になる
- モデルの仮定(前提条件)が満たされない場合には、結果が頑健ではなくなる (e.g., 初期値を変えると解も変わる)
- (明らかに正規分布に従わないであろう歪んだデータにも適用しているのが現状)
- 混合軌跡モデリング (Group-based trajectory analysis; Nagin, 2005)の紹介



# 混合軌跡モデリング

- Nagin (1999, *Psych Methods*)で整備された方法が、SASパッケージ (proc traj) として実用化されたため、ここ数年で、使用頻度が急増

proc traj論文本数推移





# 混合軌跡モデリング

## Group-Based Modeling of Development

DANIEL S. NAGIN

<http://www.andrew.cmu.edu/user/bjones/>

The screenshot shows the homepage of the 'traj' software. At the top, there is a navigation bar with links for 'home', 'download', 'examples', 'documentation', and 'contact'. The main content area features the title 'traj' in a large font, followed by the subtitle 'group-based modeling of longitudinal data'. To the left of the main text is a vertical sidebar with links for 'home', 'download', 'examples', 'documentation', and 'contact'. The main text describes the software's use of a discrete mixture model for longitudinal data and lists supported distributions: censored (or regular) normal, zero inflated (or regular) Poisson, and Bernoulli distributions (logistic model). Below this, there is a section for 'linked reference material' with two links: 'A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories' and 'Advances in Group-Based Trajectory Modeling and a SAS Procedure for Estimating Them'. At the bottom of the page, there is a copyright notice: 'Copyright (c) 2010 Bobby L. Jones'.



# 混合軌跡モデリング

- 発達軌跡にはいくつかのパターンがあるはず
- しかし、通常は、どのようなパターン（グループ）があるのかは観察できない
- クラスを仮定して、適合度比較を行うことによって、その軌跡パターンの数を考える
- ある時点の説明変数によって、軌跡への所属確率が説明される、独立変数付きのモデルも想定可能
- 3時点以上の縦断データを用いて、問題行動・認知能力などの発達や老化の様子を、グループごとに描き出し、それらの規定要因についても検討可能



## 混合成長曲線モデルと混合軌跡モデル

- 混合成長曲線モデル (growth mixture) は、潜在クラス分析と潜在成長曲線モデルを組み合わせた (ような) もの
- 潜在軌跡モデリング (proc traj) も、イメージ的には、それに類するものの、異なる分析手法
- 混合成長曲線モデルは、個人レベルの式で誤差項を仮定しているのに対して、潜在軌跡モデリングは誤差項を仮定していない (i.e., 個人のランダム効果を考えず、誤差分布を仮定しない)



## 混成成長曲線モデルと混合軌跡モデル

- 潜在軌跡モデリング (proc traj) は、群間レベルの式はパラメトリックだが、群内 (個人) レベルの式はノンパラメトリック
- よって、Naginはこのモデルをセミパラメトリックと呼ぶこともある
- 潜在軌跡モデリング (proc traj) は、時系列情報を有効に取り込んだmodel-basedなクラスタ分析と呼べるかもしれない
- (←通常のクラスタ分析は探索的な分析であるに対して)



## 結果と考察

- 抑うつ・不安症状は3クラスに分類される  
（“高・低”という2群への群分けはこのサンプルの場合は適切ではないかもしれない）
- 抑うつ・不安高群，は抑うつ・不安低群と比較した際に，気質の3次元によってそれぞれ特徴づけられる（相関と同様のパターン）
- 抑うつ・不安に対してリスクのある特定のグループを確認したり，予防的な介入のためのターゲットの絞り込みを行う際に示唆を与える





# 結果の解釈: 5つのポイント

1. 選択された群の数はいくつか?
2. 各群の軌跡の形状は?
3. 各群に属する標本のサイズや割合は?
4. 各個体の各群への所属確率は?
5. 説明変数がある場合には, それらが群分けに対してどのように寄与しているか?



# 標本サイズについて

- 混合軌跡モデリングにおいて，頑健な推定を保証するためには， $N = 300 \sim 500$ 以上の標本サイズが推奨される (Nagin, 2005)
- ただし， $N = 100$ 程度の標本サイズがあれば，分析そのものは可能
- 標本サイズが小さい場合には，検出力が小さくなり，群の数を少なく推定するおそれがある (e.g., Andruff et al., 2009)



# 混合軌跡モデリング

- 発達や老化の「類型化」を行うことがgrowth mixtureよりも統計的に頑健に可能（と言われている）
- 通常の潜在成長曲線モデル（HLM）の，さらに一歩先を行く分析（と言って間違いないはず）
- 特徴的な発達軌跡を描くグループが発見できると興味深さは増す
- 発達軌跡への所属確率を従属変数にする独立変数付きのモデルを考えることが可能
- 逆に，発達軌跡への所属確率を独立変数にする従属変数付きのモデルも構築可能
- 大規模かつ長期にわたる縦断調査データである必要はないので，3時点以上のデータには適用する価値あり



# 混合軌跡モデリング (特徴的な発達軌跡とは何か?)

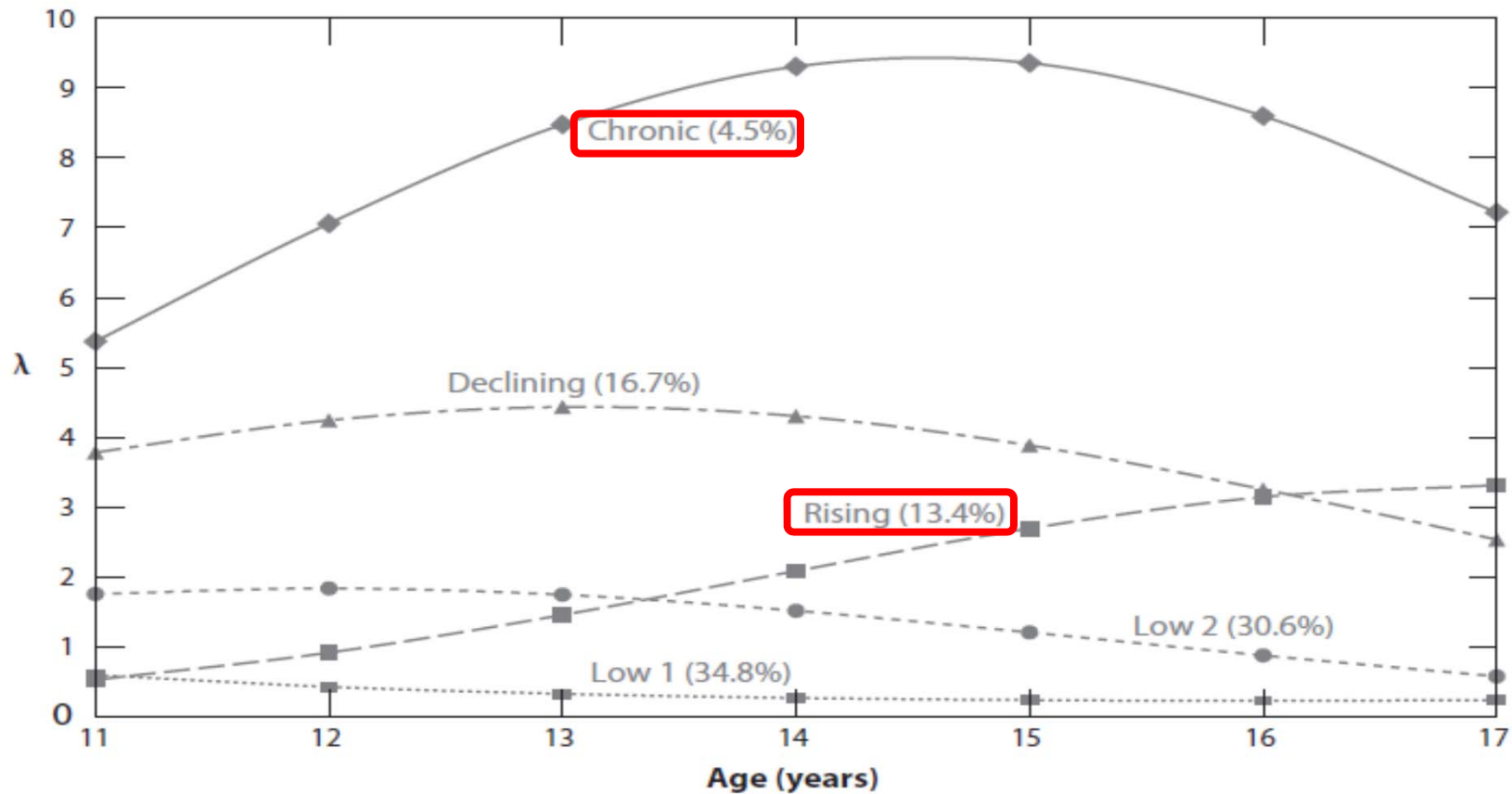


Figure 6

Trajectories of violent delinquency in the Montreal-based longitudinal study. (Data from Nagin & Tremblay 1999.)



# 実際の分析プロトコル SAS proc trajの場合

- SAS proc trajパッケージのインストールが必要
- データは、通常のDATAステップを使って読み込む
- 分析も、ほぼ通常のPROCステップに同じ

```
PROC TRAJ DATA=dataset1 OUTPLOT=OP  
OUTSTAT=OS OUT=OF OUTEST=OE ITDETAIL;  
  ID ID; VAR V1-V20; INDEP T1-T5;  
  MODEL LOGIT; NGROUPS 2; ORDER 3 3;  
  RISK X Y Z  
RUN;  
%TRAJPLOT(OP,OS,'グラフのタイトル', 'グラフの  
サブタイトル', '縦軸の名前', '横軸の名前')
```



# 実際の分析プロトコル SAS proc trajの場合

- PROC TRAJの後には各種出力データセット名を記述
- IDは、参加者IDを表す変数を指定
- VARは、従属変数を指定
- INDEPは、測定時点の変数を指定
- MODELは、従属変数の確率モデルを指定 (CNORM / ZIP / LOGITなど)
- NGROUPSは、群の数を指定
- ORDERは、各群の軌跡の次数を指定
- RISKは、独立変数(群分けを説明する変数)を指定



# 実際の分析プロトコル R CrimCVパッケージの場合

- Rパッケージ(CrimCV)も登場
- <http://cran.r-project.org/web/packages/crimCV/>
- <http://www.probability.ca/jeff/ftpdire/cross.pdf>
- 特徴は、クラス数決定のための適合度指標として、AICやBICだけではなく、Cross Validation Error (CVE)を計算してくれる点
- ただし、総当たりの行的に行うので、N数に応じて少しだけ時間を要する
- 独立変数込みのモデルは推定不可能なので、段階推定するしかない(と思われる)



# 実際の分析プロトコル R CrimCVパッケージの場合

```
library(crimCV)
```

```
load("ファイルの名前と形式")
```

```
out1 <- crimCV(ファイルの名前, ②, dpolyp=③,  
rcv=TRUE)
```

```
plot(out1)
```

```
summary(out1)
```

[群の数を指定]

[各群の軌跡の次数を指定]



2013年 10月 12日 (土) 9:30~11:30  
日本パーソナリティ心理学会第22回大会 @江戸川大学  
広報委員会企画シンポジウム

パーソナリティ心理学における統計分析の動向

発達や変化の軌跡をモデリングする

高橋 雄介

京都大学 大学院教育学研究科  
デザイン学大学院連携プログラム

E-mail: takahashi.yusuke.3n@kyoto-u.ac.jp

