

令和元年6月15日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K17305

研究課題名(和文) 縦断データ分析におけるモデルの誤設定に頑健な手法の開発と教育・発達心理学への応用

研究課題名(英文) Developing a robust method to model misspecification in longitudinal data analysis: Applications for educational and developmental psychology.

研究代表者

宇佐美 慧 (Usami, Satoshi)

東京大学・高大接続研究開発センター・准教授

研究者番号：20735394

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：複数の個人から継続的に測定したデータを縦断データと呼ぶ。縦断データを用いることで、測定値がどのように時間的に変化したかや変化の個人差、さらには変化の相関・因果関係を検証できる。本申請課題では、1)ランダム時間効果(あるイベントが全ての個人に影響し、データの従属性を生じさせるもの)を考慮した縦断データ分析法の提案、2)変数間の相関・因果関係を推論するための統計モデルの統一的枠組の提案、3)縦断データに基づく対象の分類の一手法である、構造方程式モデル決定木(SEM Tree)の性能評価検証および頑健なモデル設定法の提案、4)教育・発達心理学の縦断データへの提案手法の応用、の4つの研究課題を進めた。

研究成果の学術的意義や社会的意義

縦断データは仮説検証上の利点が多いことから、心理学、経済学・教育学・社会学・医学等多様な分野から注目を集めている。縦断データを分析するための方法論として、潜在成長モデル、クロスラグモデル、SEM Treeと呼ばれる手法等がこれまで広く利用されてきた。本研究は、実際の縦断データ分析においてその影響が生じる可能性が非常に高いにも拘わらず世界的に見ても検証が不十分であった、モデルの誤設定の問題に焦点を当てている。実践性が高い縦断データ分析手法における方法論的な問題を検証する本申請課題の遂行は、従来の理論・応用研究の双方に大きなインパクトを与えるものである。

研究成果の概要(英文)：Longitudinal design is useful because it enables researchers to effectively evaluate the trajectories of growths/changes in individuals and its individual differences. In this research project, we have proceeded the following research and published papers: (1) proposing a new statistical model that can account for time-specific effects, (2) proposing a unified framework to clarify the conceptual and mathematical relations among cross-lagged models to evaluate longitudinal associations between variables, (3) evaluating the performance of classification of individuals in SEM Trees through large scale simulation study, and (4) applying proposed methods to data in educational and developmental psychology.

研究分野：心理統計学

キーワード：縦断データ 分類 因果推論 発達 成長 回帰木 決定木 潜在成長モデル

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

小学生を対象に毎年実施して得た学校保健調査データのように、複数の個人から継続的に測定して得たデータを縦断データと呼ぶ。縦断データでは同じ個人を追跡するため、例えば各年の睡眠時間の長さの変化のパターン(傾向)や、その個人差の大きさ(e.g., 睡眠時間の長さが一定の児童もいれば、次第に短くなる児童もいる)、ならびにその規定因(e.g., 男の子の方が睡眠時間が短くなりやすい)についての研究仮説を検証できる。加えて、変数間の変化の因果関係(e.g., 睡眠時間の低下→抑うつ症状の悪化、という因果はあるが、逆の関係は成り立たない)を検証する上でも、調査・観察研究に比べより正確な推論が可能になる。このような利点から、教育・発達心理学を中心とした縦断データの収集、および分析手法の開発が世界各地で進み、日本でも縦断データ分析の専門書が刊行されるようになった。変化のパターンやその個人差を調べる方法として潜在成長モデルと呼ばれる統計モデルが、また変数の変化の因果関係を調べる為にクロスラグモデルが現在世界的に最も広く利用されている。またビッグデータの時代に対応して、変化のパターンが異なる個人を分類すると同時にその規定因をデータから探索する手法として構造方程式モデリング決定木が有望視されている。これらの統計モデルでは、それぞれ分析者が設定した様々な数学的仮定の下で成り立っている。ところが、これら一連の仮定は実際には完全には充たされない、つまりモデルの誤設定(model misspecification)が生じている事が指摘されてきた。モデルの誤設定の影響は大きく、研究仮説を検証する上で誤った結論を導く可能性を大きく高めることが近年の研究で明らかになってきた。上記の統計モデルは世界的に広く利用されており、またモデルの誤設定の影響を指摘した論文の大半は海外の一流誌で報告されていることから、モデルの誤設定の問題は高い注目を集めている。

2. 研究の目的

既存の分析法では、モデルの誤設定の影響により研究仮説について誤った結論を導く可能性が高いことが近年示されてきた。本申請課題では、縦断データ分析手法におけるモデルの誤設定の問題に関する以下の研究を行う。1) ランダム時間効果を考慮した縦断データの分析法の提案。2) 変数間の変化の因果関係を推論するための一般化クロスラグモデルの提案。3) スパコンを用いた構造方程式モデリング決定木の性能評価検証と、頑健なモデル設定法の提案。4) 教育・発達心理学の縦断データへの提案手法の応用とモデルの誤設定への一般的対処法の提案。

3. 研究の方法

1) ランダム時間効果を考慮した縦断データ分析法の提案

潜在成長モデルにおいて、存在するランダム時間効果を無視することが、タイプ I エラーをどの程度高めるのか、また効果の存在を仮定した提案モデルによってどの程度改善するのかを、理論値とシミュレーション値の双方から比較検討する。ランダム時間効果の大きさ以外にも、測定時点数、仮定する変化のパターンの形状、その他のモデルの仮定の違い等を操作しながら検証する。加えて、ランダム時間効果の影響下でもそれを適切に対処するための統計モデルを提案しその有効性を検討する。

2) 変数間の変化の因果関係を推論するための一般化クロスラグモデルの提案

変数間の変化の因果関係を推論するための統計モデルには幾つかの種類があるが、本研究では、これら既存のモデルが、(1)測定誤差を仮定しているか否か、(2)累積的な残差を仮定しているか否か、(3)個人間差を特性因子で表現しているか否か、の3つの観点から分類できることを数理的に示す。また、3つの観点を全て考慮した、一般化クロスラグモデルを提案すると同時に、既存のモデルが提案モデルの特別な場合に相等することを示す。さらに、提案モデルと既存のモデルにおける統計的性質(e.g., 因果関係を表現する母数の推定値の標準誤差の大きさ)についても比較する。

3) SEMTree の性能評価検証とモデル設定法の提案

モデルが仮定する変化のパターンの形状(e.g., 線形, 二次), グループ間の母数の等質性, ランダム時間効果の有無, の3種類のモデルの誤設定を想定する。そして、それぞれの誤設定において、真の分類(グループ)構造の再現精度の既定因として、サンプルサイズ, 測定時点数, 独立変数の測定信頼性, テンプレートモデルの複雑性(母数の数), 独立変数の数, 真のグループ数の計6つの実験条件を操作(各条件の水準数を5とし, 直交計画で計56=15,625条件を設定)して大規模なシミュレーションデータを発生させ、これら実験条件とSEMTreeの分類精度の関係を調べる。また、テンプレートモデルとして、特定の変化のパターンの形状を仮定しない統計モデルを考え、上記と同様の実験条件の下でSEMTreeの分類精度を調べ、モデルの誤設定の影響に対して頑健なテンプレートモデルの設定法を提案する。

4) 教育・発達心理学の縦断データへの提案手法の応用とモデルの誤設定への一般的対処法の提案

全国高齢者縦断調査(6000名, 約30年間に亘る縦断データ), 中高生の精神保健調査(1000名以上, 6年間), 東京ティーンコホート調査(約15,000世帯)の大規模縦断調査データを用いて研究1-3の提案手法の有効性を示す。

4 . 研究成果

上記の研究計画に則って研究を推進し , 各研究課題について英語での論文文化を行った .

5 . 主な発表論文等

[雑誌論文] (計 5 件)

1. Usami,S., Todo,N., & Murayama,K. (in press). Cross-lagged panel model in medical research: A cautionary note. PLOS ONE.
<https://www.biorxiv.org/content/early/2018/12/03/486217>
2. Usami,S., Murayama,K., & Hamaker,E.L. (in press). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*.
3. Usami,S., & Murayama,K. (2018). Time-specific errors in growth curve modeling: Type-1 error inflation and a possible solution with mixed-effects models. *Multivariate Behavioral Research*, 53, 876-897.
4. Usami,S., Jacobucci,R., & Hayes,T. (2018). Performance of latent growth model based structural equation model trees to uncover population heterogeneity in growth trajectories. *Computational Statistics*.
5. Usami,S., Hayes,T., & McArdle,J.J. (2017). Fitting structural equation model trees and latent growth curve mixture models in longitudinal designs: The influence of model misspecification in estimating the number of classes. *Structural Equation Modeling*. 24, 585-598.

[学会発表] (計 5 件)

1. Usami, S., Murayama, K. (2018). Random time errors in growth curve model. Paper presented at the 83rd Annual Meeting of Psychometric Society. New York, USA.
2. Usami, S., Murayama, K., & Hamaker, E.L. (2018). On the mathematical and conceptual relationship between cross-lagged longitudinal models. 行動計量学会第 46 回大会 慶應義塾大学 9 月
3. Usami, S., Murayama, K., Hamaker, E.L. (2017). A unified framework of cross-Lagged longitudinal models. Paper presented at the 82nd Annual Meeting of Psychometric Society. Zurich, Switzerland.
4. Usami, S., Murayama, K., & Hamaker, E.L. (2017). A unified framework of cross-Lagged longitudinal models. 統計関連学会連合大会 南山大学 9 月
5. Usami,S. & Murayama,K. (2017). Random time errors in growth curve modeling. 行動計量学会第 45 回大会 静岡県立大学 9 月

[図書] (計 0 件)

[産業財産権]

出願状況 (計 0 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
出願年 :
国内外の別 :

取得状況 (計 0 件)

名称 :
発明者 :
権利者 :

種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等
<http://satoshiusami.com/>

6．研究組織

(1)研究分担者

研究分担者氏名：なし

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2)研究協力者

研究協力者氏名：なし

ローマ字氏名：なし

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。