

# 教学IRデータにおける多変量解析手法の検討 ～他分野におけるデータマイニング先行事例との比較を通じて～

○光永 悠彦(島根大学 教育・学生支援機構 教学企画IR室)

## 教学IRデータとは

- 入試成績(科目ごと・種目別)
- 志望動機
- 授業評価アンケート
- 履修状況
- 出席状況(講義ごと)
- 履修成績(講義ごと)
- 単位取得状況(講義ごと)
- 留学履歴
- 退学・留年・休学履歴

大学における学生の履歴を、原則1行に一人の学生IDとなるようにまとめたデータ

学生ID	入試成績			履修状況			授業評価			...
	センター	2次	面接	外国語	専門	...	外国語	専門	...	
E101001	480	5	A	1			3			
E101002	500	6	A	1	1		5	3		
:										

「講義ごと・科目ごと」・・・**ネスト構造・階層データが多い**

## 多変量解析のためのデータ属性の検討

### MNAR (Missing not at random) となっている

- 例・「初級外国語を履修していなければ上級外国語を履修できない」
- 例・「選択必修科目の履修の有無」・・・背後に共通の履修トレンド要因を仮定できる
- MNARをもたらず「無視できない要因」による分析結果のバイアスの恐れ

**年度単位で完結したデータ構造** → 留年者が同一講義を2度受講した場合等、例外多し

尺度水準・・・  
**名義尺度** 履修科目等  
**順序尺度** 志望順等  
**間隔尺度以上** 入試成績, 授業評価, 出席回数等, 多数

→ **尺度水準が混在している**

データマイニング手法を取り入れることで、教学IRデータに対しより適切に解釈できるモデルを構築できる可能性がある

**目的**

教育分野以外におけるデータマイニング事例で用いられている因果推論等の方法論を教学IRデータに援用する可能性を検討

## マーケティング分析手法の教学IRデータへの適用

マーケティングで用いられるデータマイニング手法例(岡太・守口, 2010を改変)

ニューラルネットワーク・・・売上予測, 顧客ターゲットング, 優良顧客の識別  
 MDS, コレスポネンス分析・・・ポジショニング分析, 消費者セグメント分析  
 多項ロジットモデル・・・ブランド選択行動分析, 店舗選択行動分析  
 決定木・・・顧客ターゲットング, 優良顧客の識別  
 アソシエーションルール・・・商品の併売行動分析

**教学IRデータの分析**

顧客→「学生」  
 消費者→「学生」  
 店舗→「講義」  
 併売行動→「同時受講」等  
 置き換えて分析を実施

**教学IRデータ分析のためのモデルの精緻化と適用可能性の拡張**

## データマイニングで用いられるモデルの精緻化, 適用可能性の拡張

データ融合(POSデータの分析)(星野, 2009, p.194)

欠測値への対処(多重代入法+階層モデル)

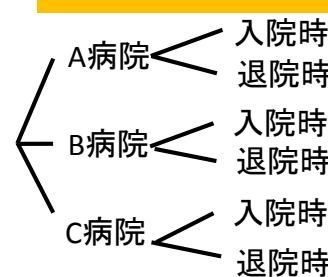
互いに異なる集団

購買履歴データ	欠測
欠測	市場調査データ

**共変量**

同一人物からデータが得られない場合, 同一属性を持つであろう別の集団からのサンプルに共通して現れる**共変量**を用いてデータを「つなぐ」  
 → **傾向スコア**による因果推定, **欠測部分の推測**

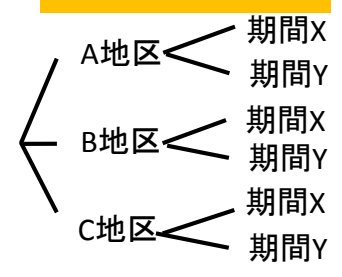
入退院履歴データ



**欠測データ**

疫学・医療統計＝ランダムサンプリングが困難な場合  
 POSデータ＝地区ごとの特徴を平準化したい場合

地区別POSデータ



**欠測データ**

## 教学IRデータの分析における応用可能性

X学部学生 Y学部学生

X学部専門科目成績	欠測
欠測	Y学部専門科目成績

**共変量(共通教育科目他成績)**

**予測**

**予測**

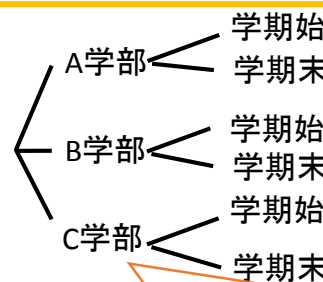
→ 傾向スコアによる因果推定, 欠測部分の推測＝**仮にX学部の学生がY学部のみに開講されている科目を受講した場合の成績を予測できる**

【その他のメリット】

- ・ 専門科目間での「単位取得難易度」の標準化が可能

【課題】

- ・ 共変量の選定によって結果が大きく変わる可能性
- ・ 予測結果の一般化が限定される(年度によって異なる結果が得られた場合の解釈)



**欠測データ**

**完全データ**

**因果モデルのあてはめ**

学生の属性, 学部の属性, 授業科目の配当, 期別, ...

階層のネスト・・・「**変量効果**」と考える

→ E-Mアルゴリズムによるパラメタの推定(Swaminathan & Rogers, 2008)を, 多重代入法を用いて推定した完全データに対して行う

【課題】

- ・ **学部の差異**を特徴づける分析は, 慎重に行われなければならない
- ・ 学生の階層化は, クラスタリングの妥当性の検討を要する

## 解決すべき課題

【階層内に人数が少ない場合の推定法】

- ・ 階層の併合等, 推定が可能となるクラスタ数を探索・・・恣意的な決定とならないよう留意・・・変量効果とみなして意味のある階層か? 要検討

【カテゴリカルデータに対する潜在因子を仮定した分析】

- ・ IRT(item response theory)による方法の検討・・・欠測データにも対応できるが, 1次元性が仮定されなければならない。また, 項目特性の解釈に工夫が必要

【参考文献】 Swaminathan, H. & Rogers, H.J. (2008). Estimation procedures for hierarchical linear models. In O'Connell, A. & McCoach, D.B. (eds). *Multilevel Modeling of Educational Data*. Charlotte, NC: Information Age Publishing.

星野崇宏(2009). 調査観察データの統計科学. 岩波書店 / 岡太彬訓・守口剛(2010). マーケティングのデータ分析. 朝倉書店