

MCAの基本形から応用へ

計量分析セミナー2025

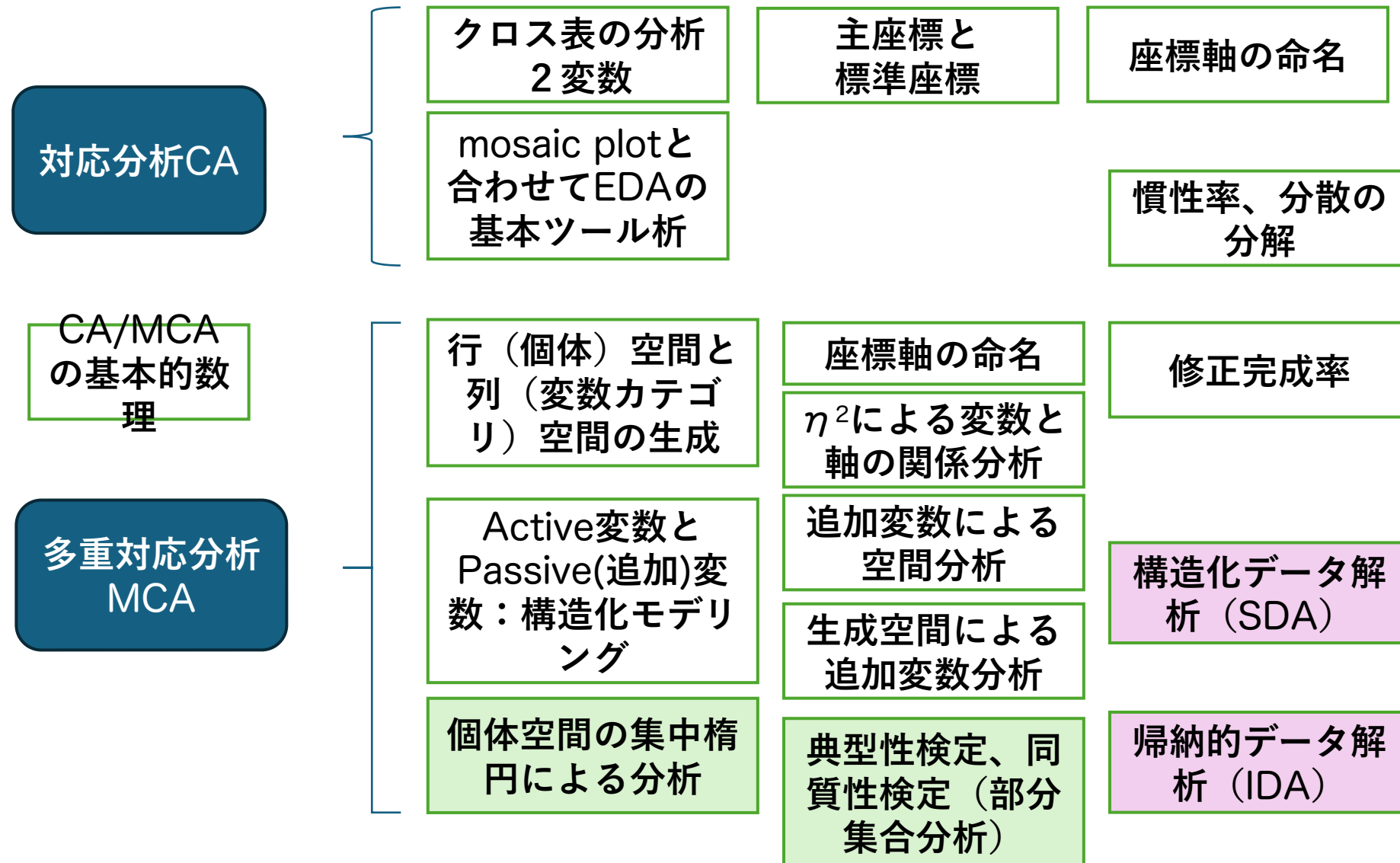
2025年9月4日

津田塾大学 数学・計算機科学研究所

藤本一男

kazuo.fujimoto2007@gmail.com

このセミナーでお話しした内容です



■ は時間の関係で説明できない可能性あり。資料は用意してあります。

MCAの基本機能

- 入力
 - 行に個体、列に変数の調査集計表
- MCA
 - 行はそのまま、列を変数カテゴリの指示行列（indicator行列）に。
 - それに対してCA
 - この処理は、MCAのアプリの中でやってくれるので、意識する必要はない。
 - 低頻度カテゴリには、注意が必要。junk指定する。
- 通常の調査データだと、NA、無回答など、低頻度カテゴリがあって、それが、グラフをひっぱる。
 - それらをjunk指定してMCAを行う、specific MCAという拡張がオーソドックスなMCA。

データの前処理のポイント

- NAは、文字列で“NA”や”無回答”に変換しておくこと。
- 変数は、因子（factor）に変換しておくこと。
- カイ二乗検定はしておくこと（CA）
- 度数分布表を作成して変数の分布を確認しておくこと。

MCAの本番

- 名称こそ、specific がついているが、なにか「特別な」処理ではなく、調査集計表の分析の際にはほぼ必須の機能。
 - FactoMineR::MCAでもexcl=でjunkカテゴリの番号を指定可能。
 - GDAtools::speMCAは、excl=NULLのままだと、普通のMCA。
- MCAの最初の作業は「慣性率」の確認。
 - GDAtools::speMCAの結果では、res.speMCA <- speMCA(...)としてたら
 - res.speMCA\$eig で参照できる。
 - 表で判断だけでなく、グラフで表示するほうがよい。

構造化設計

- 空間を生成する変数の選択
 - 生成する座標軸は、この変数のカテゴリからの合成変数になるので、合成が不可能なものは選んではいけない。
 - 数学的には、座標軸はでてくるが、解釈が困難！
- そこに射影する変数の選択

変数カテゴリ、個体の分布を見る

- `GDAtools::ggcloud_variables(res.speMCA)`
- `GDAtools::ggcloud_individuals(res.speMCA)`

で変数カテゴリと個体の分布を確認する

- junkカテゴリを指定しないと、大抵、低頻度カテゴリが大きく引張っている図が描かれることになる。そのカテゴリを`excl=`で指定してやって、再度、`speMCA`を行います。
 - junkカテゴリの番号を取得する方法
- `ENQview::Shiny_speMCA` を使うと動的に変化をみることができる。

junkを除去（excl）しないと

- exclしないと、分析したいポイントがグシャ！と固まってしまっしまい分析できない…。
- junkをexclしてから、再度、speMCAを行うと、分析対象をひろげて表示することができる。
- junk指定するカテゴリは、よく吟味すること。低頻度であっても、重要なカテゴリはある。

ここで、最初の課題

- 生成された座標軸は、個体マップでも変数マップでも同じ。
- この生成された座標軸は「新たな変数」です。
- 最初の課題は、この変数に名前をつけること。
 - ブルデューの「ディスタンクシオン」での社会空間の変数マップでは、
 - 縦軸が「資本総量」、横軸が、+経済資本/-文化資本、と-経済資本/+文化資本、と命名されている。
 - これは、投入する変数に依存。
 - 数学的には、なにか座標軸は生成されるけれど、それが解釈可能（もしくは期待しているもの）になるかどうかは別の問題。
 - 軸に対する変数の寄与を勘案して、軸を命名することになる。

どうするのか

- 軸に対する変数カテゴリの大きさと向きを表にして確認する。
- 重要度の大きいカテゴリを睨んで、えい！と命名する。
 - 重要さは、カテゴリの寄与率としてあたえられているので、生成AIに投げ込んで、いくつかの候補をだしてみても、考察するのもありだと思う。

ここで、もう一つ！追加変数という変数

- 冒頭、構造化設計で触れたところ。
- 空間を生成する変数は、Active変数と呼ばれる。
- この変数は、均質なものでないといけない。合成した状態を考えると、困難なカテゴリをまぜてしまうと、解釈ができない。
- しかし！
- 年齢/年代、性別、などのような回答者属性で、その空間をの構造を分析したいことはよくある。
- ここで使えるのが、追加変数と遷移公式。
- 遷移公式は、行と列の変数が、内的に連関していることを数式で表現したもの。03 CAの数理(2) を参照

ある時期まで変数空間の分析までだった

- 変数カテゴリマップがある。
 - それで座標軸の解釈をおこなうと、空間生成に寄与した変数間の関係を分析できる。
 - 加えて、追加変数を使うことで、空間生成に寄与した変数とは異質な変数をつかって、分析が可能になる。
 - 変数カテゴリポイントは、他方の個体空間のすべてのポイントの総和が反映している。（逆もなりたっている。個体ポイントには、変数カテゴリのすべてのポイントの総和が反映している。）
- ここで、どの軸に、どの変数のばらつきが表現されているか（たとえば、性別の違いは第3軸、というように）を見ることができる。相関比2乗 (η^2)

つづけて、個体マップの分析方法が

- 個体マップでは、各回答者は、平均値などに抽象化されずに、それぞれの位置を維持している。
- この位置は、回答した変数カテゴリによって、ばらつきもっている。
 - 変数の平均位置を個体マップの中に定めることはできるが、ばらつきは、別のこと。
 - このばらつきを表現するのに、集中楕円を描くこと。
 - この集中楕円は、その変数カテゴリを選択した個体のひろがり、楕円で表現するもので、その楕円は、そのカテゴリを選択した個体の86%が含まれているもの、として近似的に表現している。

この集中楕円の分析で検定を行う

- 見た目、楕円がずれているとする。
- ならば、分布は同じではない、という解釈になるのか。
- 平均位置と分散をつかって検定することがここでも行われる。
- 典型性検定
 - ある分布の位置が、全体の分布と同じなのか違うのかを検定。
- 同質性検定
 - 二つの楕円の位置と分布を違うものとして解釈していいか、否かを検定。
- これらは、IDA（帰納的データ解析）と呼ばれる。
- MCA2021や、スライド「A05_SDAからIDAへV2.0」を参照

いろいろあるが、最も重要な機能

- それは、Active変数によって、空間を生成し、それに追加変数を射影することによって空間の解釈を行う、という仕組み。
- 逆に、追加変数を生成された空間という「定規」でもって、解釈を行うという仕組み。
- これは、参照する「入れ物」「定規」を明示的に設定する分析方法であるということ。
- この入れ物として均質なものを想定できる工場内の製品の状態などであればともかく、社会のように各要素が複雑な関係をもっているところでは、その複雑な状態を複雑なまま前提にする方法が必要。

つづき

- そうした空間を前提にして、回帰分析や分散分析を行えばよいのでは。
 - そうした参照空間を明示しない分析手法は、均質な参照空間を前提にしているはず。
 - 個体空間の表現の一つは、集中楕円であるが、個体空間事態の密度表示も行われる。
 - そうした濃淡のある空間を前提に、追加変数の解釈も可能なのである。
 - ANOVA（分散分析）で、実験計画をおこなって、実際に存在する環境のばらつきなどをデータの処理でNULL化（均一化）できるようにするが、社会調査の場合は、そのような方法での均一化はできないので、個体の分布密度などを考慮することで、補正しようとする。