

“集まるデータ”を活かすには…

質的データの マイニングのための対応分析法

質的データのきめ細かい分析の“コツ”や非定型・非構造化データに有効な分析法を学ぶ

● 講座の概要

対応分析法(コレスポネンス分析、数量化法Ⅲ類)は、“質的データの分析”に適した多次元データ解析手法です。マーケティング・リサーチや社会調査などで多用する選択肢型質問や自由回答質問などで扱うデータの多くは“質的データ”です。しかもデータ取得環境の多様化に伴い、データの形態は複雑になる傾向にあります。最近は、さまざまなデータが大量かつ容易に“集まる”時代といわれております。ビッグデータ・アナリティクスやソーシャル・メディアで扱うデータも、量的データだけではなく質的データが混在し、それがこうしたデータの特徴でもあります。

分析技法の自動化もかなり進んでおりますが、このことはそのままデータ解析の信頼性にはつながりません。“データは集まる”時代であると同時に、自分の利用目的に合った質のよいデータを的確にみきわめて“集めること”が必要です。集めるための知恵が必要で、調査設計や実験計画の役割も看過できません。確かなデータにもとづき、きめ細かい分析を行う“コツ”を知ることがデータ科学の本質です。

統計ソフトウェアや分析ツールも、利用環境の変化に合わせて、また長年の多くの研究成果を取り入れることで、多彩で複雑な機能を実装するようになっております。対応分析法も大抵のソフトに組み入れられさまざまな機能が利用できますが、これが分析過程の暗箱化につながり、誤用や濫用を生むという弊害もみられます。

対応分析法は、成熟した完成度の高い、よく知られた手法です。テキスト型データはもとより、さまざまな形態(非定型、非構造化、半構造化)のデータセットの分析・探査に役立つことから、その利用価値が再認識されている“古くて新しい”手法です。

対応分析法はフランスで誕生し、おもに欧州圏で進展をみた方法論です。また数量化法Ⅲ類は、日本で誕生し国内で独自の展開をみた数量化理論の中の一手法です。いまでこそ、両者は数理的には同等手法であることは周知のことですが、両者の思想的な背景には、かなり異なるものがあります。とくに対応分析法は、フランス語圏の文献・資料が多く、独自の概念や用語を用いることもあって、その本質が正確に理解されてきたとはいえない面もあります。本講座では講師が独自に用意した豊富な資料を用いて、対応分析法の本来の考え方を忠実に辿りながら、平易に解説いたします。また、手法の特性を活かした、テキスト型データの解析に適した大規模で疎なデータ表の対応分析法やクラスター化法についても取り上げます。これらの基本的な考え方、“適切な使い方”、“べきべからず”を、数値例や事例分析を用いて紹介いたします。

具体的には、次ページの「講座内容」に挙げた諸要素について説明いたします。また、典型的な統計ソフトウェアである JMP® (ジャンプ) や WordMiner® (ワードマイナー) を用いた試用体験の演習日を設けました(希望者のみ)。

● キーワード

対応分析法 (CA : Correspondence Analysis/AFC: Analyse Factorielle des Correspondances)、数量化法Ⅲ類 (パターン分類)、2元クロス表と2元データ表の分析、クロス表の独立性の検定とカイニ乗統計量、多重対応分析法 (MCA : Multiple Correspondence Analysis)、多重クロス表 (パート表)、アイテム・カテゴリー型データ、インジケータ行列、クラスター化法 (ハイブリッド法)、非定型・非構造化データ、自由回答・自由記述、テキスト型データ、テキスト・マイニング、大規模で疎なデータセットの処理、双対性、プロファイルとその同等性、成分スコア (数量化スコア)、カイニ乗距離とユークリッド距離、布置図・同時布置図と視覚化

● 日程及びプログラム

■ 全3回+ 総合演習日 (ソフトを用いた演習を含む)

第1回 2016年11月30日(水) 13:00~17:00

第2回 2016年12月7日(水) 13:00~17:00

第3回 2016年12月14日(水) 13:00~17:00

総合演習日(自由参加)2016年12月20日(火)13:00~18:00

(*) いずれの日も、小休憩時間(コーヒー付)を用意いたします。

講座内容

(※) 以下の各項目は、この講座で扱う予定の内容で、厳密にこの順に説明を行うものではありません。また、数理的な細かい話しはなるべく避け、基本的な情報について丁寧に説明いたします。

第1回 2016年11月30日(水) 13:00~17:00

● 質的データの数量化とは

- まえおき：帰納的、仮説発見的なアプローチの重要性
- データの種類、データ表の形式を知ること
- 質的データとは(名義尺度、順序尺度)
- 量的データ(区間尺度、比例尺度)との違い
- その他のデータ(数値・非数値、構造化・非構造化など)
- なぜ質的データの“数量化”が必要か

● 変量・項目間の連関性とグラフィカル・ツールによる探査

- 多変量構造のデータと基本となるデータ表
- クロス表の意味と重要性、項目間の連関性の意味
- クロス表の探査・分析に必要な情報
- クロス表のモザイク図による連関性の観察
- 多重クロス表(パート表、パート行列)とその特徴
- 多重クロス表と多変量モザイク図
- 視覚化の効用と限界

● 変量・項目間の連関性の測度 - 変量間の関連を測る -

- 連関性の測度とは
- ピアソンのカイ二乗統計量の特徴
- 2元クロス表の独立性の検定の意味
- 対応分析におけるピアソンのカイ二乗統計量の役割

第2回 2016年12月7日(水) 13:00~17:00

● 対応分析(CA: correspondence analysis)の基本的な考え方

- 対応分析と数量化Ⅲ類、その他の類似手法(要約)
- 対応分析(CA)とは(その基本原理)
- 2元データ表への拡張(対応分析法で扱うデータ表)
- プロファイルとプロファイル空間(重心座標系とは)
- 対応分析における合成変数の作り方と考え方
- 成分の解釈と成分スコアの主な性質(双対性、分布の同等性ほか)
- カイ二乗距離とユークリッド距離
- ピアソンのカイ二乗統計量との関係(慣性・総変動と固有値・特異値など)
- 成分スコアの配置図、同時配置図の解釈
- 配置図による視覚化の限界、適切な利用法(よくある誤用の回避)

第3回 2016年12月14日(水) 13:00~17:00

● 質的データのクラスター化法

- 対応分析法とクラスター化(自動分類)
- 対応分析との併用の効用
- 対応分析におけるクラスター化の特徴
- データ表の行と列の同時分類
- 成分スコアのクラスター化とその留意点
- 大規模データのクラスター化(ハイブリッド法の適用)

● テキスト・マイニング: テキスト型データの分析への応用

- なぜテキスト型データに適用可能なのか
- テキスト型データとその特徴(要約)
- 疎な大規模データセットの分析とその特徴
- 簡単な分析例

● 多重対応分析(MCA)とは

- 多重対応分析(MCA)の考え方
- 多重クロス表(パート表・パート行列)とクロス表の関係
- 数量化Ⅲ類(アイテム・カテゴリー型、インジケータ行列)との関係
- 大規模データ表の分析への応用 - 追加処理の応用 -

● 対応分析、多重対応分析の適用例

- 簡単な例、トイ・データによる観察
- 社会調査データ、ウェブ調査データの分析例
- 集計でえた2元表の分析(一般の集約データ表)
- 多重クロス表の分析、その他

演習日(自由参加) 2016年12月20日(火) 13:00~18:00

ソフトを用いた演習形式で行います。ミニチュアデータを用いて、手法の仕組みを知り分析手順と結果解釈の要点を学ぶこと、実際データの分析に必要なシナリオ策定を模擬し、これにそって分析体験することが目標です。

- 統計ソフトウェア JMP® およびテキスト・マイニング・ツール WordMiner® を用いた演習を行います。
- ご自身のPCをお持ちいただけます。
- 貸出用PCをご希望の方は、**台数に限りがございますので、事前にご連絡くださいますようお願いいたします。**
- サンプル・データセットをご用意いたします。もちろん、ご自分でご用意のデータセットを用いた分析を行うことも可能です。
- 参加者の方が、ご自分のPCにJMP® およびWordMiner® のトライアル版をインストールする方法や演習の要領については、講座開催時に別途ご案内いたします。

● 講師

大 隅 昇 統計数理研究所・名誉教授

講師プロフィール

文部科学省統計数理研究所・調査実験解析研究系助教授、教授を経て、情報・システム研究機構・統計数理研究所名誉教授(2004年から)：博士(理学)・専門社会調査士、日本計算機統計学会ソフトウェア賞受賞(テキスト・マイニング・ソフト WordMiner の開発他)

専門分野: データ科学、多次元データ解析、調査方法論、テキスト・マイニングなど

主要著書: 「統計的データ解析とソフトウェア」(日本放送出版協会)、「記述的多変量解析法」(日科技連出版社) / 「調査法ハンドブック」(監訳)(朝倉書店) / 「社会調査ハンドブック」, 「多変量解析実例ハンドブック」などの分担執筆、その他

ホームページ: <http://wordminer.org/>

● 対 象

- いままでに、対応分析などを利用して、結果解釈に悩んだり、仕組みが分からずに困った人
- 質的データの分析、テキスト・マイニングなどに関心のある人、使ってみたい人
- 市場調査／マーケティング・リサーチ分野のリサーチャー、データ分析者・データアナリスト
- 統計科学の初等的な知識があることが望ましいが必須条件ではない（ただし、平均値、分散と標準偏差、相関係数、カイ二乗統計量などの記述的統計量の意味が分かっていると具合がよい）
- 線形代数の基礎知識があることが望ましいが必須条件ではない（行列とベクトル、できれば固有値問題など）

● 定 員

定員：30名

※10名に達しない場合は、開催を延期させていただく場合がございます。

● 会場および連絡先

一般社団法人 日本マーケティング・リサーチ協会
〒101-0044 東京都千代田区鍛冶町1-9-9 石川LKビル2階
TEL 03-3256-3101 FAX 03-3256-3105

● 参加費・支払い方法

JMRA 正・賛助会員社（者）／1名

42,000円＋消費税＝ 45,360円

一 般／1名

60,000円＋消費税＝ 64,800円

振込先 三菱東京UFJ銀行 飯田橋支店(普通) 4669542
みずほ銀行 飯田橋支店(普通) 660769
(銀行振込時に振込手数料は差し引かずをお願いします。)

- 請求書と参加証は10月31日より発送を開始いたします。
- 請求書到着後、11月29日(火)までにお振込みください。
- 参加者が10名に達しない場合は、開催を見合わせていただく場合がございます。その際は、事前にお申し込み者にご連絡申し上げます。

● お 願 い

参加費はご欠席されても返金できません。お申し込みされた方がご都合の悪い場合は代理の方がご出席ください。代理の方で参加も難しい場合は、7日前までにご連絡ください。それ以降の場合は下記のキャンセル料を申し受けますので予めご了承ください。

キャンセル料

7日前～2日前のキャンセル …………… 参加費の20%
前日のキャンセル …………… 参加費の70%
当日キャンセル、ご連絡がなかった場合 …………… 参加費の100%
また、当日の録音機器等の持ち込みは、固くお断り致します。

<個人情報取り扱いについて>

下記の個人情報の取扱いに関する事項についてご確認いただき、同意の上お申し込み下さい。

- 事業者の氏名または名称 一般社団法人 日本マーケティング・リサーチ協会
- 個人情報保護管理者 一般社団法人 日本マーケティング・リサーチ協会 事務局長
- 個人情報の利用目的 取得した個人情報は、お申込み受付対応およびセミナー運営のために利用いたします。「登録の可否」欄に「可」としていただいた場合は、上記に加えて、JMRA メールニュース（月に一度の業界の情報をお届けするメルマガ）や参加セミナーの関連情報を配信するために利用いたします。
- 個人情報の第三者提供について 本人の同意がある場合または法令に基づく場合を除き、取得した個人情報を第三者に提供することはありません。
- 個人情報の取扱いの委託について 取得した個人情報の全部または一部を委託する場合があります。その場合には、個人情報の管理水準が、当協会が設定する基準を満たす企業等を選定し、適切な管理、監督を行います。
- 開示対象個人情報の開示等および問合せ窓口について 本人からの求めにより、当協会が本件により取得した開示対象個人情報の利用目的の通知・開示・内容の訂正・追加または削除・利用の停止・消去（「開示等」といいます。）に応じます。開示等に応じる窓口は、下記になります。
一般社団法人 日本マーケティング・リサーチ協会
個人情報 苦情・相談窓口責任者 (TEL: 03-3256-3101 FAX: 03-3256-3105)

き り と り



一般社団法人 日本マーケティング・リサーチ協会主催

《2016年度 質的データのマイニングのための対応分析講座》

1. JMRA 正会員 2. JMRA 賛助法人会員 3. JMRA 賛助個人会員 4. 一般

会社名

所在地 〒

TEL

FAX

お名前	役職名 (部・課名)	E-mail	統計分析業務 経験年数	登録の 可(※1) 否(※2)
			年	可 否
			年	可 否

※上記登録の可否をいずれかに、○印をお願いします。

※1 参加申込書に記入された個人情報は、JMRAの各種イベント、各種講座・会合のご案内、書籍等の販売、その他の関連情報等をお知らせする書面、Eメール、電話、FAX等の連絡に使用することを同意いたします。

※2 今回のセミナーにおける参加確認以外に使用する事には同意しません。

FAX 03-3256-3105 お問合せは TEL: 03-3256-3101
WEBからのお申込みは <http://www.jmra-net.or.jp/seminar/>

JMRA 一般社団法人
日本マーケティング・リサーチ協会