

京都女子大学データサイエンス研究所
令和5年度事業報告

令和6年12月
京都女子大学データサイエンス研究所

目次

概要		1
1	企業・自治体との連携	2
2	データサイエンスブレイクスルー事業	4
2.1	第1回データサイエンス統計領域シンポジウム	4
2.2	第1回データサイエンス情報領域シンポジウム	6
3	データサイエンス実践教育・研究の推進事業	9
3.1	ダッソー・システムズ株式会社との連携事業	9
3.2	2023年度スポーツデータサイエンスコンペティション	11
4	データサイエンスカフェ事業	13
4.1	IT業界キャリアデザイン入門	13
5	データサイエンス学部・研究所開設記念シンポジウム	16
6	その他の活動	19
6.1	アドバイザー会議	19
付録		
	第1回データサイエンス統計領域シンポジウム講演論文集	

概要

データサイエンス研究所は、女子大学初となるデータサイエンス学部の開設に先立ち、京都女子大学の建学の精神に則り、本学のデータサイエンスに関する物的・人的・知的資源を活用して、行政・産業界・教育機関と連携し、データサイエンスによる社会の課題解決およびイノベーション創出に貢献することを目的として、令和4年4月に開設されました。

データサイエンス研究所では2つの目標を掲げており、一つ目は、データサイエンス分野に多様性の展開を図ることであり、女子大学の特性を生かしたジェンダーや価値観、ライフスタイルなどに新たな方向性を導くとともにデータサイエンスの文系分野への拡充を目指すものです。二つ目は、産業界、メディア、政策担当者など様々なステークホルダーと連携しデータサイエンスのブレークスルーを推進し、先端的な研究の進化と体系化およびデータサイエンス・AIを駆使してイノベーションを創出できる人材の育成を目指しています。

令和5年度は、4月にデータサイエンス学部が新たに開設し、研究所の構成員として学部教員が兼担研究員として加わり、2月にはデータサイエンス学部・研究所開設記念シンポジウムを開催しました。本研究所の令和5年度事業として、「研究所・センター提案型プロジェクト」への申請、「統計エキスパート人材育成プログラム」への参加を行い、企業・自治体との連携活動、社会ソリューション分野・統計分野・情報分野の3つの分野ごとにセミナー・シンポジウムの開催、学会への学生参加の支援、データサイエンスカフェ(DS Café)の整備などの事業を行いました。その他、研究所にアドバイザー制度を導入し、5名の先生方をお迎えしてアドバイザー会議を開催しました。本報告書では、これらデータサイエンス研究所令和5年度事業について報告します。

1 企業・自治体との連携

データサイエンス研究所では、データサイエンスによる社会の課題解決およびイノベーション創出に貢献することを目的に本学のデータサイエンスに関する物的・人的・知的資源を活用して、行政・企業・教育機関との連携を行っています。データサイエンス研究所・データサイエンス学部開設に伴い、本学でも企業・自治体とのデータサイエンス分野での連携協定を締結しています。令和5年度末までに下記の15の企業と連携協定の締結を行っています。

1. 日新電機株式会社【2021/2/1】
2. 株式会社 SCREEN ホールディングス【2021/8/18】
3. 京都市【2022/3/28】
4. オムロンヘルスケア株式会社【2022/3/28】
5. 京都薬科大学【2022/6/16】
6. 西日本旅客鉄道株式会社（JR 西日本）【2022/10/31】
7. 京都府【2022/11/16】
8. NEC ネットエスアイ株式会社【2022/12/13】
9. 株式会社リンクレア【2023/4/1】
10. ダッソー・システムズ株式会社【2023/9/5】
11. ダイハツ工業株式会社【2023/10/2】
12. 株式会社ナイトレイ【2023/10/2】
13. エイターヘルスケア株式会社【2023/10/2】
14. 株式会社インテージホールディングス【2023/11/22】
15. 西宮市【2023/12/21】

これら企業には、データサイエンス学部の講義として本学でのご講演、学生らの現場見学として企業・庁舎の見学等を実施させていただくなど、本学の教育活動にご協力いただきました。令和5年度に連携活動を実施した企業からは、本学への入学を志望する中

高生に向けてデータサイエンスの活用分野を広く知らせるためロゴマークをご提供いただき、データサイエンス学部の web ページ (<https://www.kyoto-wu.ac.jp/gakubu/faculty/data/renkei.html>) にて公開しています。本学と連携協定を締結していない企業で令和 5 年度末までにロゴマークを提供いただいた企業は次のとおりです。

1. 京都市産業技術研究所
2. 株式会社 JTB
3. 京セラ株式会社
4. 塩野義製薬株式会社
5. 西日本電信電話株式会社 (NTT 西日本)
6. 中外製薬株式会社

上記の企業・自治体以外にも来年度以降の連携に向けた打ち合わせなどを随時実施しており、データサイエンス研究所・データサイエンス学部を中心に京都女子大学でのデータサイエンス教育を今後も推進していきます。

2 データサイエンスブレイクスルー事業

データサイエンス研究所に所属する研究員は、(1) 数理統計分野（ヘルスサイエンス、データアナリシス等）、(2) 情報分野（AI・機械学習等）、(3) 社会ソリューション分野（地域連携、マーケティング等）のいずれか（または複数）の分野で研究を行っています。データサイエンス研究所は今年度より「研究所・センター提案型プロジェクト」に申請し、本プロジェクトの一環として、ステークホルダーと連携したデータサイエンスブレイクスルー推進事業を実施しました。

この事業では、本研究所の強みを生かした分野研究において、先端的な研究の進化と体系化を推し進めるとともにデータサイエンス・AIを駆使したイノベーションの創出を目的とし、企業・自治体・外部研究機関との連携や共同研究を進めています。また、各グループ内での研究成果を広く公表するため、研究所所属の教員およびその関係者が参加する研究集会やシンポジウムを開催しています。

令和5年度では、数理統計分野・情報分野のグループにおいてシンポジウムを開催しました。ここでは、各分野のグループで実施されたシンポジウムについて報告します。

2.1 第1回データサイエンス統計領域シンポジウム

データサイエンス研究所統計分野のグループでは、今年度「統計エキスパート人材育成プログラム」への参加を行っていることもあり、統計分野において現在も活躍されている研究者の先生方をお招きし最新の研究成果について講演をいただくため、第1回データサイエンス統計領域シンポジウムを下記の日程で開催しました。

日時：令和5年12月16日（土）9:30～17:00

場所：本学 E 校舎 E005 教室

本シンポジウムでは、全部で3つのセッションを行い、計12件の講演が行われました。各講演では、多様な分野での統計理論の活用事例などが報告され活発な議論が行われました。招待セッション I では、主にスポーツデータの分析からヘルスデータの分析について講演が行われました。特別企画セッション「データサイエンスと統計」では、統計関連のレジェンドと呼ばれる先生方をお招きし、先生方が活躍されている分野について示唆に富む総合報告をしていただきました。招待セッション II では、因果推論や空間統計に関する講演が行われました。本シンポジウムには、データサイエンス学部の学生も参加し、大いに刺激を受けていました。

当日のプログラムは次の通りです.

招待セッション I

座長：足立浩平（大阪大学）

大学スポーツ観戦調査とその分析

横山暁（青山学院大学）

測定技術の進化とデータサイエンス～野球を中心として～

小泉和之（順天堂大学）

卓球におけるデータサイエンスの活用とゼロ変形交差一般化幾何分布による打球回数のモデリング

酒折文武（中央大学）

ヘルスデータサイエンスの現状と取り組み

富田誠（横浜市立大学）

医療リアルワールドデータを用いた薬剤耐性対策における臨床疫学・薬剤疫学研究と今後の薬学教育への展開

村木優一（京都薬科大学）

特別企画セッション「データサイエンスと統計」 座長：栗原考次（京都女子大学）

臨床試験を哲学する

手良向聡（京都府立医科大学）

長崎大学 情報データ科学部のあゆみとこれから

西井龍映（長崎大学）

データサイエンスにおける時空間統計科学の役割

矢島美寛（東京大学）

データサイエンスと医学統計学 ～統計的因果推論の視点から

岩崎学（順天堂大学・統計数理研究所）

招待セッション II

座長：阿部貴行（京都女子大学）

統計的因果推論の考え方と確率的介入に基づく因果効果

山本倫生（大阪大学）

エシエロン解析法：最近の展開と応用事例

石岡文生（岡山大学）

Covid-19 新規感染者数のデータ収集と可視化

山本義郎（東海大学）



開会式の様子

2.2 第1回データサイエンス情報領域シンポジウム

データサイエンス研究所情報分野のグループでは、情報学分野で活躍されている3名の先生方をお招きして第1回データサイエンス情報領域シンポジウム「情報学のダイバーシティ AI・ロボット・宇宙」を下記の日程で開催しました。

日時：令和6年2月17日（土）13:00～17:00

場所：本学 E 校舎 E005 教室

シンポジウムでは、3名の先生から、サービスロボット、ネットワークダイナミクス、天体シミュレーションに関する講演が行われました。積極的に議論に参加し、研究に対する深い洞察を共有しあうなど、シンポジウム全体で活発な知識交流が行われ、様々な学問領域に関して活発な議論が行われました。データサイエンス学部の学生も参加し、大いに

刺激を受けていました。

当日のプログラムは次の通りです。

サービスロボティクスの実用化に向けたフロンティアと課題

エル ハフィ ロトフィ (立命館大学)

ネットワーク系の非線形ダイナミクスとコンピューティング

鈴木 秀幸 (大阪大学)

コンピュータシミュレーションで探る宇宙の構造と起源

小久保 英一郎 (国立天文台)



開会式の様子

京都女子大学 第1回 データサイエンス 情報領域シンポジウム

「情報学のダイバーシティ AI・ロボット・宇宙」

開催日：2024年2月17日(土)
 場所：京都女子大学 E校舎
 時間：13:00～17:00
 (12:30受付開始)

※ Bing Image Creatorにより作成

情報学はデータサイエンスの主要な領域であり、その対象分野は多様です。このシンポジウムでは、計算、ロボット、天文学など様々な情報学の分野において第一線でご活躍されている講師陣を迎え、最新の研究成果についてご講演いただきます。皆様のふるってのご参加をお待ちしています。

**「サービスロボティクスの実用化に向けた
フロンティアと課題」**

講師：エル ハフィ ロトフィ 先生
(立命館大学)



**「ネットワーク系の非線形ダイナミクス
とコンピューティング」**

講師：鈴木 秀幸 先生
(大阪大学)



**「コンピュータシミュレーションで探る
宇宙の構造と起源」**

講師：小久保 英一郎 先生
(国立天文台)



お申し込み方法：

参加費：無料
 専用フォームからお申し込み下さい



<https://forms.office.com/r/Y59CSf3BiF>

 **京都女子大学** データサイエンス学部
 データサイエンス研究所

第1回データサイエンス情報領域シンポジウムのポスター

3 データサイエンス実践教育・研究の推進事業

データサイエンス関連の学会や企業・自治体などが主催するコンペティションやハッカソンなどの様々な企画に本学の学生の参加を促し、活動の支援を行うためのデータサイエンス実践教育・研究の推進事業を実施しています。令和5年度は、本学と連携協定を締結しているダッソー・システムズ株式会社との共同事業として、令和5年度総務省地域デジタル基盤活用推進事業「スマート道路灯実証事業」に参画し、データサイエンス学部の学生を対象としたワークショップを本学にて開催し、実証実験が行われている静岡県裾野市にてワークショップの成果についてプレゼンテーションを行うため有志学生の派遣を行いました。また、令和6年1月6日から7日まで中央大学後楽園キャンパスにて日本統計学会が開催した2023年度スポーツデータサイエンスコンペティションに有志の学生5名の参加を支援しました。ここでは、ダッソー・システムズ株式会社との連携事業と2023年度スポーツデータコンペティションへの参加について報告します。

3.1 ダッソー・システムズ株式会社との連携事業

本学と令和5年度に連携協定を締結したダッソー・システムズ株式会社との連携事業としてデータサイエンス学部と協力し、令和5年度総務省地域デジタル基盤活用推進事業「スマート道路灯実証事業」に参画しました。本事業は、NTTコミュニケーションズ株式会社を代表とし、ダッソー・システムズ株式会社を含む複数の企業が参画して、裾野市でスマート道路灯の実証実験を行っています。

ワークショップの実施

ダッソー・システムズ株式会社の連携事業として、データサイエンス学部1回生を対象とした「スマート道路灯実証事業」に関するワークショップを次の日程にて開催しました。

日時：令和5年12月12日（火）14:30～17:15

場所：本学 図書館

参加者：学生54名、企業9名

ワークショップでは、参加者を学生7～8名と企業の社員1名で構成される8グループに分け、スマート道路灯の活用法・発展性などに関するブレインストーミングを行い、得られたアイデアを整理しました。アイデア出しの作業は各30分の3つのセッションに分け

て行われました。

セッション	テーマ
1st セッション	スマート道路灯で何を表示し、伝え、どんな効果を期待するか。
2nd セッション	スマート道路灯でどんな情報やデータを集め、どんなことに活かすか。
3rd セッション	スマート道路灯をどのようなヒト・モノ・コトと組み合わせると、より活用できるか。どんな新しい使い方があるか。

各セッション 30 分と短い時間で実施されましたが、非常に多くのアイデアが提示されました。主なアイデアとしては、AI センサーやクラウドサービス、情報通信インフラに接続して緊急度の高い災害対策や犯罪防止に注目したもの、利用者の性別・年齢などの情報を収集し、ビジネスや観光産業に活用するもの、学生らしいアイデアも多数提示され、社会人ではなかなか気づかない視点から斬新なアイデアとの評価もいただきました。本ワークショップは、実証実験が行われているスマート道路灯について具体的なサービスへとブラッシュアップしていく取り組みへと繋がる有意義なものとなりました。

静岡県裾野市役所でのプレゼンテーション

ワークショップでグループごとに提示されたアイデアについて、ワークショップ参加者からさらに希望者を募集し、静岡県裾野市役所にてプレゼンテーションを行うため学生の派遣を行いました。プレゼンテーションは次の日程で実施されました。

日時：令和 6 年 2 月 1 日（木）15:00 ～ 16:00

場所：静岡県裾野市役所

参加者：学生 18 名

プレゼンテーション発表に際し、グループごとに約 5 分程度のプレゼンテーションを実施しました。各グループの発表では、スマート道路灯の路面に文字や画像を表示する機能とローカル 5G を生かし、交通事故の防止・道路の渋滞の解消といった交通課題の解決、災害時の情報収集や避難所への案内、動機検知センサーによる不審者の検知といった防犯対策など、さまざまな活用案が示されました。また、裾野市内で開催されるイベントでのスポンサー表示、猪や鹿といった害獣による被害への対策などユニークなアイデアも提案されました。発表終了後には、裾野市の村田はるかぜ市長より、学生による柔軟なアイデ

アを高く評価していただき、本活動を通じて身に着けた AI などに関する知識と問題解決を行う能力について引き続き伸ばしてほしいと、学生に対する激励の言葉をいただきました。

当日の会場には、市役所職員の皆様、スマート道路灯実証事業に参画する企業の皆様、地元メディアの記者の皆様が出席されており、学生らの発表の様子は静岡新聞をはじめ複数のメディアにて取り上げられました。

3.2 2023 年度スポーツデータサイエンスコンペティション

令和 6 年 1 月 6 日～7 日に中央大学後楽園キャンパスにて日本統計学会が開催する「2023 年度スポーツデータサイエンスコンペティション」インフォグラフィック部門にデータサイエンス学部 1 回生の有志の学生 5 名が参加しました。学生らは、2020 年度から 2022 年度までの 3 年間の日本プロ野球の各選手のデータのうち、特に身長と体重に着目し BMI を算出、「BMI25 未満」、「BMI25 以上 30 未満」、「BMI30 以上」の 3 つの層での層別分析を行い、それぞれの層で打者や投手の特徴を比較しました。学生らは、分析結果についてまとめたポスターを作成し、コンペティションの期間中、常時オンライン上にて公開されました。

参加した学生らは、会場やオンライン上で行われた口頭発表やポスター発表にも参加し、スポーツデータのクリーニング法や解析方法などについて深く学びました。口頭発表やインフォグラフィック部門以外のポスター発表では、データサイエンス学部の講義等で学んだ手法も頻繁に登場しており、学生らも興味をもって学んでいる様子でした。会場では学外の研究者や大学院生と積極的に議論を行い学術的な交流も図られました。

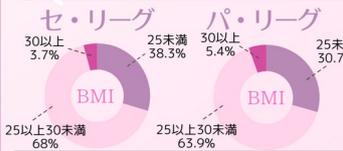
ヘルスケアから野球を診る



【診たいこと】

- ◆各リーグにおける**球団の成績**（順位）と**BMI**には何らかの傾向があるのか？
- ◆打者や投手の項目とBMIに何らかの**パターン**や**特徴**をみることができるのか？

点1 リーグと順位との関係



【説明】
 ※ 筋力の発達によりBMIは高くなるが、野球以外のスポーツ選手のBMIからし、順位は3順位と推察される。
 ※ したがって、本研究では日本野球学会のBMI年齢¹⁾に基づきBMI25や30で階別しグラフを作成した。
 ※ 順位データがない、2022年度のデータはエラー発生により数値がなかったためリーグ別のBMIの比較に
 関わるグラフ作成はすべて分析対象から除外した。
 ※ 盗塁数と死球数は、各選手の平均値をBMIで階別し代打数（平均値）により集計した。

＜参考文献＞
 1) URL: http://www.jpbc-sports.jp/journal/column/042010/column_018.htm (公益財団法人日本野球学会
 雑誌「ベースボール」) 2010年1月号
 2) URL: http://www.japbc.or.jp/data/magazine/pdf/mediamagazine2022_2.pdf (日本野球学会 雑誌「野球」2022年
 2月号) 2022年2月号
 3) 2021年度日本プロ野球選手会 選手会報告書 (資料番号) 2022.060606
 4) 高田文彦・藤原真樹・佐藤 誠・内藤正樹・保科和也(2017). 野球のドラフト指名データに基づいた1対1の投手対戦率
 偏りの分析. 日本体育科学, 45(2), 107-112.
 5) Slawik JS, Aung KT, O'Brien AZ, Cain EJ, Dugas JR, Fleiss CS. (2019). Fastball Velocity and Elbow-Volar Torque in
 Professional Baseball Pitchers. Journal of Athletic Training, 54(3), 296-302.
 6) Prodromo J, Patel N, Kainer N, Dorethy K, Tabb LP, Tom J. (2016). Pitch Characteristics Before Ulnar Collateral Ligament
 Reconstruction in Major League Pitchers Compared With Age-Matched Controls. Orthopaedic Journal of Sports Medicine,
 4(3), 23259716653896.
 7) Microsoft Corporation. Microsoft Corporation 利用規約およびその他の利用に関する登録商標または商標
 8) Microsoft PowerBI. Microsoft Corporation の利用規約およびその他の利用に関する登録商標または商標
 9) R Core Team (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing,
 Vienna, Austria. <https://www.R-project.org/> (バージョン 4.2.2)
 10) https://www.canva.com/ja_jp/learn/branding/#/branding/size (ブランドサイズ) 参照

点2 打者・投手の項目との関係

BMI=体重÷身長²

	盗塁数		球種	
	2021年	2022年	2021年	2022年
BMI25未満	0.035	0.035	BMI25未満	92.2
BMI25以上 30未満	0.0094	0.0092	BMI25以上 30未満	86.6
BMI30以上	0.0021	0.0045	BMI30以上	83.3

	死球		球速	
	2021年	2022年	2021年	2022年
BMI25未満	0.0034	0.0049	BMI25未満	44.7
BMI25以上 30未満	0.0098	0.0062	BMI25以上 30未満	47.3
BMI30以上	0.017	0.018	BMI30以上	75.0

【診れたこと】

BMIが低い
 BMIの平均値が低い
 チームの成績が上がる(?)
 盗塁が多い ⇒ 高い敏捷性により転倒を中心、死球が多い ⇒ ケガのリスクが高まる(?)
 としてのケガの予防につながる(?) 直球系の速度が速い ⇒ 肩や腕にかかる負担
 が多くケガをしやす(?)

BMIが高い
 BMIの平均値が高い
 チームの成績が下がる(?)
 球種が多い ⇒ 肩や腕にかかる負担が
 少なくケガをしにくい(?)

※ 作品制作に用いたソフトウェアは、Excel[®]、PowerPoint[®]、R[®]、Canva[®]である。

学生らが作成したポスター：黒塗り部分は学生氏名

4 データサイエンスカフェ事業

データサイエンス研究所では、データサイエンスに関するソーシャルラーニングスペースとしてデータサイエンスカフェ（DS Café）を設置し、その整備を行うためデータサイエンスカフェ事業を行っています。具体的には、本研究所のデータサイエンスの拠点化という目標に向けて、より充実したものとするため、(1) 資格取得（統計検定など）の勉強会、(2) 外部から講師を迎えた定期的なセミナー、(3) 関係者との交流会のようなイベントを学生有志の協力のもと企画・開催しています。また、データサイエンスの拠点化を目指すにあたり、裾野広げることも重要であるため、(4) データサイエンスに関する啓発活動のためのイベントの開催を実施し、中高生をはじめとして、データサイエンスの理解を広める活動の実施を目指しています。

令和5年度は、DS Caféの運用開始にあたり学生と教員の交流が可能なソーシャルラーニングスペースとしての整備を行いました。主には統計分野、情報分野、社会ソリューション分野のグループから教員が数名分担し、学生らが講義内容の質問や勉強方法の相談などが可能な時間を設けました。前期は主に水曜日、後期は主に火曜日に実施しました。

(2) に示した外部からの講師によるセミナーとして、社会ソリューション分野のグループによって本学と連携協定を締結している株式会社リンクレアの協力のもと、全3回のIT業界のキャリアデザイン入門を実施しました。ここでは、株式会社リンクレアのミニ講義について報告します。

4.1 IT業界キャリアデザイン入門

データサイエンス研究所の社会ソリューション分野のグループと株式会社リンクレアの協力のもとで、データサイエンス学部1回生を対象とした「IT業界キャリアデザイン入門」を全3回のミニ講義を本学にて実施しました。

本講義では、株式会社リンクレアで現在活躍中の社員の皆様に講師としてお招きして、仕事に関するリアルなお話をしていただき、IT業界およびIT企業で働くという事について学生へ向けにご講演いただきました。データサイエンス分野での女性人材として、実際のロールモデルである女性社員の皆様の話を聞くことで学生らも自身の将来のキャリア像についてイメージをすることができました。講義後には、講師と若手社員の皆様との懇談会も実施していただき、学生らも和やかな雰囲気の中でコミュニケーションを取ることができました。

各講義の概要は次の通りです.

第1回：IT 業界, IT 企業のリアル

日時：令和5年10月25日(水) 14:45～16:15

場所：本学 E 校舎 E101 教室

講師：吉沢 均 (株式会社リンクレア 代表取締役社長), 森高 寛子

参加者：40 名

第2回：システムエンジニア (SE) の仕事

日時：令和5年11月15日(水) 14:45～16:15

場所：本学 E 校舎 E101 教室

講師：長岡 牧子, 馬谷 幸恵

参加者：40 名

第3回：IT コンサルタントの仕事

日時：令和5年12月6日(水) 14:45～16:15

場所：本学 E 校舎 E101 教室

講師：樋口 潤子, 今井 佑

参加者：7 名

データサイエンス学部生向け

《IT業界》 キャリアデザイン 入門<全3回>



この講座では、(株)リンクレア(独立系SIer、創業53年)で現在活躍中の方々から仕事に関する“リアル”なお話を聞き、IT業界およびIT企業で「働く」ということについて理解を深めます。また、さまざまな年代・役割を担う女性の働き方を知ること、自分自身の将来「なりたい姿・キャリア」について考えるきっかけとしてください。

講義の後には講師や若手社員との『懇談会』を予定しています。リラックスした雰囲気の中で、疑問点など質問することができます。奮ってご参加ください。

第1回 IT業界、IT企業のリアル

2023年10月25日(水) 14:45~16:15 E101教室

<講師>吉澤 均(株式会社リンクレア 代表取締役社長)、森高 寛子(経営企画室、1996年入社)

<内容>IT業界が社会をどのように支えているか、変えているのか。IT企業が担う社会の中の役割、仕事内容について。企業内の各部署がどのような仕事をしているのか。

第2回 システムエンジニア(SE)の仕事

2023年11月15日(水) 14:45~16:15 E101教室

<講師>長岡 牧子(SE、1997年入社 京都女子大学卒)、馬谷 幸恵(SE、2008年入社)

<内容>情報システム構築の全体の流れ・工程を説明し、その工程におけるシステムエンジニアの役割、仕事内容について実際の事例をもとにイメージを深める。

第3回 ITコンサルタントの仕事

2023年12月06日(水) 14:45~16:15 E101教室

<講師>樋口 潤子(取締役、2006年中途入社 京都女子大学卒)、今井 佑(ITコンサルタント、2011年入社)

<内容>情報システム構築におけるITコンサルタントの役割、仕事内容について、実際の事例をもとに理解を深める。複雑性・不確実性が高まる時代に必要となる新たな「コンサルテーション」について紹介する。

皆さんがデータサイエンス学部で学び始めて半年が過ぎました。大学での学びが実際に社会でどう役立つのか、どう活かされるのか、是非そのつながりを実感してください！

IT 業界キャリアデザイン入門ポスター

5 データサイエンス学部・研究所開設記念シンポジウム

令和4年度のデータサイエンス研究所の開設，令和5年度のデータサイエンス学部の開設を記念し，データサイエンス学部・研究所開設記念シンポジウム「データサイエンス教育の今，そしてこれから」を次の通り開催しました。

日時：令和6年2月3日（土）13:00～17:00

場所：ホテルグランヴィア京都3階「源氏の間」

本シンポジウムでは，文部科学省，京都府，京都市からの来賓挨拶，滋賀大学 竹村彰通 学長からの記念講演に続き，京都女子大学におけるデータサイエンス教育・研究の現状について，連携先機関の皆様から事例発表を行いました。

「京都女子大学データサイエンス学部の教育と研究」の報告では，データサイエンス学部1回生から，2023年度スポーツデータサイエンスコンペティションへ参加し発表をした報告，総務省令和5年度地域デジタル基盤活用推進事業 AI搭載スマート道路灯プロジェクトの一環で静岡県裾野市にて発表をした報告が行われました。

パネルディスカッションでは「データサイエンス系の学部に何が必要か？何を期待するか？」「企業・自治体との連携のあり方」「女性データサイエンティストへの期待」「京都女子大学データサイエンス学部への期待」等について議論していただき，データサイエンス教育の今とこれからの考える機会となりました。

当日のプログラムは次の通りです。

開会挨拶

竹安 栄子（京都女子大学 学長）

来賓挨拶

藤澤 亘（文部科学省研究振興局学術基盤整備室 室長）

山下 晃正（京都府 副知事）

尾崎 学（京都市 デジタル化戦略監）

記念講演

竹村 彰通（滋賀大学 学長）

報告「京都女子大学データサイエンス学部の教育と研究」

栗原 考次（京都女子大学 データサイエンス学部長）

柴田 敏之（京都薬科大学 研究科長）

京都府 総合政策環境部大学政策課

京都市 総合企画局情報化推進室

熊野 和久（ダッソー・システムズ株式会社 パブリックセンターディレクター）

北西 由武（塩野義製薬株式会社 DX 推進本部データサイエンス部長）

京都女子大学データサイエンス学部 学生

パネルディスカッション

竹村 彰通（滋賀大学 学長）

椿 広計（統計数理研究所 所長）

山本 章博（京都大学データ科学イノベーション教育研究センター長）

小野 陽子（横浜市立大学データサイエンス学部 准教授, WiDS アンバサダー）

佐伯 諭（データサイエンティスト協会 事務局長）



竹安学長による開会の挨拶の様子



栗原学部長による報告の様子

京都女子大学 データサイエンス学部・研究所 開設記念シンポジウム

開催日時：令和6年2月3日(土) 13:00~17:00
開催場所：ホテルグランヴィア京都 3階「源氏の間」
(京都市下京区烏丸通塩小路下ル JR 京都駅中央口直結)

※ Bing Image Creatorにより作成

開会挨拶	竹安栄子 (京都女子大学 学長)
来賓挨拶	藤澤亘 (文部科学省研究振興局学術基盤整備室 室長) 山下晃正 (京都府 副知事) 尾崎学 (京都市 デジタル化戦略監)
基調講演	竹村彰通 (滋賀大学 学長)
報告 「京都女子大学データサイエンス学部の教育と研究」	
栗原考次 (京都女子大学 データサイエンス学部長) 柴田敏之 (京都薬科大学 研究科長) 京都府 (総合政策環境部大学政策課) 京都市 (総合企画局情報化推進室) 熊野和久 (ダッソー・システムズ(株) パブリックセンターディレクター) 北西由武 (塩野義製薬(株) DX推進本部データサイエンス部長) 京都女子大学データサイエンス学部学生	
パネルディスカッション	
竹村彰通 (滋賀大学 学長) 椿広計 (統計数理研究所 所長) 山本章博 (京都大学データ科学イノベーション教育研究センター長) 小野陽子 (横浜市立大学データサイエンス学部 准教授、WiDSアンバサダー) 佐伯諭 (データサイエンティスト (DS) 協会 事務局長)	
お申し込み方法	下記専用フォームからお申し込み下さい (定員になり次第受付を終了いたします) https://forms.office.com/r/KC3bm22nE0
	  京都女子大学 データサイエンス学部・研究所

データサイエンス学部・研究所開設記念シンポジウムポスター

6 その他の活動

6.1 アドバイザー会議

今年度よりデータサイエンス学部が新たに開設され、企業・自治体との連携活動、データサイエンスに関わる女性人材育成のための教育活動が本格化したことを踏まえ、データサイエンス研究所にもアドバイザー制度を導入しました。本学のアドバイザー制度の規則に従い、次の5名の先生方をアドバイザーとして迎えました。

- 竹村 彰通（滋賀大学 学長）
- 椿 広計（統計数理研究所 所長）
- 山本 章博（京都大学データ科学イノベーション教育研究センター長）
- 小野 陽子（横浜市立大学データサイエンス学部 准教授，WiDS アンバサダー）
- 佐伯 諭（データサイエンティスト協会 事務局長）

5名のアドバイザーの先生方と研究所研究員が参加した令和5年度アドバイザー会議を次の通り実施しました。

日時：令和6年2月2日（金）14:00～16:00

場所：本学 L 校舎 会議室 1

アドバイザー会議では、令和5年度の本研究所およびデータサイエンス学部の取組について統計分野、情報分野、社会ソリューション分野から報告を行い、これら研究所・学部の活動についてアドバイザーの先生方からご意見をいただきました。各教員の研究活動に加え、今後研究所・学部の活動が増えることを踏まえて教員の負担が増加することへの対応、来年度以降のデータサイエンス学部への入学者確保の方策など課題点もご指摘いただきました。アドバイザー会議当日の議題は次の通りです。

1. 京都女子大学データサイエンス研究所について
 - (1) 2023年度活動報告について
 - (2) 産官学連携事業について
 - (3) 統計エキスパート人材プロジェクトについて
 - (4) その他

2. 京都女子大学データサイエンス学部について

- (1) データサイエンス学部における教育について
- (2) 企業・自治体と連携した教育について
- (3) データサイエンスカフェについて
- (4) 総務省地域デジタル基盤推進事業への参加について
- (5) 日本統計学会 2023 年度スポーツデータサイエンスコンペティションへの参加について
- (6) その他

3. 意見交換

付録

第1回データサイエンス統計領域シンポジウム
講演予稿集

京都女子大学データサイエンス研究所

2023年12月16日(土)

京都女子大学

京都女子大学データサイエンス研究所・第1回データサイエンス統計領域シンポジウム
於 京都女子大学 E 校舎
〒605-0981 京都府京都市東山区今熊野北日吉町 35

12月16日(土)

09:30～ 受付開始

10:00～10:10 開会

10:10～11:50 招待セッション I

座長：山田隆行（京都女子大学）

大学スポーツ観戦調査とその分析 . . . 1

横山暁（青山学院大学）

測定技術の進化とデータサイエンス～野球を中心として～ . . . 5

小泉和之（順天堂大学）

卓球におけるデータサイエンスの活用とゼロ変形交差一般化幾何分布による
打球回数のモデリング . . . 7

酒折文武（中央大学）

ヘルスデータサイエンスの現状と取り組み . . . 11

富田誠（横浜市立大学）

医療リアルワールドデータを用いた薬剤耐性対策における
臨床疫学・薬剤疫学研究と今後の薬学教育への展開 . . . 13

村木優一（京都薬科大学）

11:50～13:20 昼休み

13:20～15:20 特別企画セッション「データサイエンスと統計」

座長：栗原考次（京都女子大学）

臨床試験を哲学する . . . 17

手良向聡（京都府立医科大学）

長崎大学 情報データ科学部のあゆみとこれから . . . 19

西井龍映（長崎大学）

データサイエンスにおける時空間統計科学の役割 . . . 21

矢島美寛（東京大学）

データサイエンスと医学統計学 ～統計的因果推論の視点から～ . . . 23

岩崎学（順天堂大学・統計数理研究所）

15:20 ~ 15:40 休憩

15:40 ~ 16:40 招待セッションⅡ

座長：阿部貴行（京都女子大学）

統計的因果推論の考え方と確率的介入に基づく因果効果 25

山本倫生（大阪大学）

エシエロン解析法：最近の展開と応用事例 27

石岡文生（岡山大学）

Covid-19 新規感染者数のデータ収集と可視化 31

山本義郎（東海大学）

16:40 ~ 17:00 総合討論

17:00 閉会

18:30 ~ 情報交換会

本シンポジウムは統計エキスパート人材育成コンソーシアム（京都女子大学）との共催で開催いたします。

大学スポーツ観戦調査とその分析

青山学院大学 横山 暁

0 はじめに

本稿および発表は、スポーツ庁が募集した「大学スポーツ振興の推進事業」に対し2017年度に青山学院大学が応募し採択された際に実施した調査研究(青山学院大学編, 2017), および, 継続調査として2018年度と2020年度に青山学院大学経営学部グローバルビジネス研究所の研究として, 2019年度に青山学院大学総合プロジェクト研究所の「スポーツマネジメント研究所」の研究として実施した一連の調査研究(横山, 2019;2020;2021a;2021b;2021c;2022; 横山・山下・中邨・芳賀, 2019)を基にしたものである。

1 研究の概要

大学スポーツの調査, とりわけ観戦者に対する調査研究は林(2014, 2017)があり, 観戦経験や再観戦の意向, また大学への帰属意識に関する研究がなされている。また帰属意識に関する研究においては二宮(2011)などに見られる。これらの研究における調査は, 試合会場で観戦者に対するアンケートや, チームのウェブサイト訪問者に対してのアンケートという形を取っている。

上に記した一連の研究では, 調査会社のモニタに対してウェブアンケート形式で観戦経験の有無を含めてアンケートを実施した。2018年度の調査では, プロスポーツおよび大学体育会のスポーツの観戦経験や意向, また大学への帰属意識を測る質問を設定した。調査期間は2019年2月23日~28日, 調査対象は関東・関西・東海の大学の62の大学に所属するに調査を実施した。また2019年度の調査では, プロスポーツおよび大学体育会のスポーツの観戦経験や意向に加え, 上記既存研究(林, 2014;2017; 二宮, 2011)を参考に大学への帰属意識の設問をより詳細に設定した。調査は2020年2月18日~24日, 調査対象は関東の23の大学の在学学生および卒業生と関西の16の大学の在学学生に対して実施した。それぞれの調査の詳細等は横山(2021, 2022)を参照のこと。なお, 2020年度も2021年2月に同様の調査を実施したが, 2020年度は新型コロナウイルス感染症の流行により, 大学スポーツを含む多くのスポーツイベントの開催が行われなかった・無観客での開催であったため, 本研究では分析の対象外としている。

2 基礎集計

2.1 2018年度の調査

2018年度の調査の分析では, 1266名の回答者のうち, 体育会に所属している学生を除いた910名を対象として分析を実施した。この調査では, 過去3年間におけるプロスポー

ツの試合会場での観戦経験や同行者、観戦理由、TVでの観戦経験・頻度と観戦理由、また非観戦者の観戦したい競技、特にBリーグの観戦経験や意向、また大学の体育会の観戦経験、同行者、観戦理由、体育会の各スポーツへの関心、観戦意向について調査している。本稿では横山(2021)3節を基に大学スポーツに関する調査部分の集計結果を簡単に紹介する。なお発表では詳細な分析結果および考察について述べる予定である。

大学スポーツの観戦経験を問う質問に対しては50.5%の学生がいずれか1つの競技での観戦経験があると回答した。また、競技ごとの観戦経験をみると、駅伝や野球は2割程度の観戦経験があるのに対し、その他の多くのスポーツの観戦経験がないことが分かった。さらに関東地区の大学で調査時から見て過去5年で東京箱根間往復大学駅伝競走(以下、箱根駅伝)に出場していた大学、出場していない大学、その他の地域の大学で観戦経験を集計したところ、箱根駅伝出場大学が30.1%、非出場が26.6%、その他の地域が13.3%であった。また、東京六大学野球に所属する大学か否かでも集計したところ、六大学は33.9%、それ以外は16.5%と大きな差が見られた。

大学スポーツの観戦と大学の校歌や応援歌が歌える・知っているという質問や大学のグッズを所持しているかどうかの質問との関連性について集計したところ表1となり、それぞれ関連性が見られる結果となった。

表 1: 大学スポーツの観戦経験と校歌・応援歌および大学グッズの所持

	校歌や応援歌が歌える・知ってる	大学グッズ所持
観戦経験あり	42.6%	40.9%
観戦経験なし	30.4%	26.2%

2.2 2019年度の調査

2019年度の調査の分析では、特に関東の大学の在学学生および卒業生の分析を行った。体育会に所属している・していた在学学生・卒業生を除いたそれぞれ350名、546名を分析対象とした。本稿では前節同様に横山(2022)3節を基に集計結果を簡単に紹介する。なお発表では詳細な分析結果および考察について述べる予定である。

大学スポーツの観戦経験を問う質問に関して、少なくとも1つの競技を1回以上観戦したことのある割合を集計したところ表2となった。在学学生に関しては「過去1年における観戦経験」について調査している一方、卒業生に関しては在学中および卒業後の全期間における観戦経験について調査しているため、単純に比較できないが、在学中の1年間で2割の観戦率だとすると、卒業時にはおおむね6割程度が観戦する結果となると考えられ、在学学生と卒業生で在学中の観戦率に大きな差は見られないと判断できる。

競技ごとに集計したところ、在学学生は野球と駅伝の観戦経験の割合が高く、卒業生は野球と駅伝に加えラグビーやサッカーも比較的観戦している結果となった。なお詳細は横山(2022) p.229, 表3を参照のこと。

表 2: 在学生・卒業生の大学スポーツの観戦経験

在学生（過去1年）	卒業生（在学中）	卒業生（卒業後）	卒業生（合計）
20.0%	56.4%	37.9%	68.7%

3 重複クラスター分析法による分析

2018年度の調査では、比較的観戦経験者数の多い野球、ラグビー、駅伝、アメリカンフットボールの4つの競技に絞り、調査対象である関東および関西の合計15の大学ごとに観戦経験を平均化したデータを用いて類似度データを作成し、重複クラスター分析法 (Shepard & Arabie, 1979) を用いて大学スポーツ観戦経験から見る大学間の関係性について分析を実施した。

その結果、クラスター数5の結果が採用され、関東の大学によって構成されるクラスター、関西の大学によって構成されるクラスターが出現したとともに、東京六大学に所属する大学で構成されるクラスターも出現する結果となった。

さらに2019年度の調査に関しては、対象とする大学は関東の13の大学とし、調査で取り上げた11のスポーツの「在学生」「卒業生（在学中）」「卒業生（卒業後）」の観戦経験を個人差として個人差重複クラスター分析法 (INDCLUS; INdividual Differenced CLUStering; Carroll & Arabie, 1983) を用いて分析を実施した。

その結果、クラスター数5の結果が採用された。多くの大学が所属するクラスターが2つ現れ、卒業生（卒業後）の重みが小さいクラスターと大きいクラスターであった。よってそれぞれ一般的な学生の在学中の観戦クラスター、卒業後の観戦クラスターと解釈される。また、東京六大学に所属する大学が所属しないクラスターも出現した。特にこのクラスターは在学生、卒業生（在学中）の重みが大きく、非東京六大学の在学中の観戦傾向（観戦しない傾向）と解釈された。

これらの分析の詳細は当日紹介するが、大学間の観戦傾向の特徴を把握することが出来たと考えられる。

4 まとめと今後の課題

本研究では大学スポーツ観戦の調査に関する基礎集計や重複クラスター分析法を用いた大学間の特徴の把握について紹介した。新型コロナウイルス感染症の影響もあり、大学スポーツの観戦傾向に変化がみられている可能性も考えられる。今後も継続して調査・分析が必要であると考えられる。

また、2019年度の調査では大学の帰属意識に関する質問も設定している。大学スポーツと帰属意識の関係性を分析することで、大学スポーツの在り方について検討することも可能であると考えられ、今後の研究課題としたい。

参考文献

- [1] 青山学院大学 編 (2018). 「地方創生型大学スポーツ提案拠点の形成 成果報告書」, 平成 29 年度 スポーツ庁 大学横断的かつ競技横断的統括組織 (日本版 NCAA) 創設事業 (大学スポーツ振興の推進) . https://www.mext.go.jp/prev_sports/comp/a_menu/sports/micro_detail/_icsFiles/afieldfile/2018/06/14/1406134_001.pdf (第四章 pp. 25-75 担当)
- [2] Carroll, J. D., & Arabie, P. (1983). INDCLUS: An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika*, 48, 157-169.
- [3] 林直也 (2014). 大学アメリカンフットボールの試合観戦者における再観戦意図に影響を及ぼす観戦動機に関する研究, *身体運動文化論攷*, 13, 107-126.
- [4] 林直也 (2017). 大学への帰属意識に影響を及ぼす要因に関する研究: 大学スポーツチームに対するアイデンティティおよび観戦動機との関係に着目して, *Human welfare*, 9, 179-192.
- [5] 二宮浩彰 (2011). プロスポーツ観戦者行動におけるチームに対する愛着とホームタウンへの地域愛着, *同志社スポーツ健康科学*, 3, 14-21.
- [6] Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86, 87-123.
- [7] 横山暁 (2019). 大学生のスポーツ観戦に関する調査. *日経広告研究所報*, 53 (5, Serial No. 307), 54-59.
- [8] 横山暁 (2020). 大学スポーツの応援経験調査分析 ～大学スポーツ応援経験は何をもたらすか～. 立命館大学スポーツ健康科学研究センターシンポジウム「with コロナ/Post コロナ時代におけるスポーツの未来像—経営学・イノベーションから切り拓く—」. (2020年9月26日, 立命館大学 (オンライン開催))
- [9] 横山暁 (2021a). クラスタ分析を用いた大学スポーツ観戦意向に関する調査データの分析. *青山経営論集*, 55 (4), 187-199.
- [10] 横山暁 (2021b). 大学スポーツの観戦経験と帰属意識に関する研究 [要旨]. *日本計算機統計学会第 35 回大会講演論文集*, pp. 137-140.
- [11] 横山暁 (2021c). 大学スポーツにおける観戦者調査, 花内誠 編, 「ASC 叢書 3 スポーツビジネスの「キャズム」」 (第 7 章, pp. 181-199), 晃洋書房.
- [12] 横山暁 (2022). 重複クラスタ分析を用いた大学間の大学スポーツ観戦の分析. *青山経営論集*, 56 (4), 226-239.
- [13] 横山暁・山下勝・芳賀康浩・中邨良樹 (2019). 大学におけるスポーツに関する観戦者調査 [要旨]. *日本計算機統計学会第 33 回大会講演論文集*, pp. 59-60.

測定技術の進化とデータサイエンス ～野球を中心として～

順天堂大学 小泉 和之

1 はじめに

データサイエンスという言葉は Peter (1974) により初めて使われたと言われているが、その当時のデータサイエンスは定義も曖昧であり、学問として確立し、本格的に実用化されるまでには多くの時間を費やした。現在我々の目の前にあるデータサイエンスが大きく発展していくのはここ 20 年くらいの話になると考えられる。その発展の背景には

- 計算機の飛躍的な性能向上
- 測定技術の発展と多様化

という影響が大きいだろう。

計算機の性能向上がもたらしたものの一つとして、これまでの理論のみに頼るデータサイエンスから計算機を利用した方法論の構築が可能となったことがある。機械学習と呼ばれる方法はまさに計算機の発展と対をなすように目覚ましい発展を遂げており、一時はある機械学習の国際会議における半分以上の講演がアルゴリズムに関する研究成果の発表であったことから計算機をいかに利用して手法を発展させるかということがこの分野において大事であるかがわかる。また、機械学習のみならず、深層学習、人工知能、統計学（これらが現在データサイエンスと呼ばれている）においても計算機なしには研究や実装することは出来なくなってきている。

さらに、もう一つ挙げるとすれば、データの蓄積が容易になったことがある。例えば、これまでは紙ベースなどで保管をしていたデータが電子的に保管できるようになり、データ活用へのハードルが低くなったこともデータサイエンスの発展に大きく貢献している。データの処理でも計算機の影響は大きく、これまでは処理時間の関係で使うことが難しかった方法も現在では実用できるまでになっているという例も少なくない。

また、測定技術の進化はデータ取得の多様化をもたらした。例えば、これまでは最先端の機器でしか取得が出来なかったようなレベルのデータがスマホ 1 つで多くの高性能なデータ（画像、動画、GPS 情報など）が取得できるようになった。

2 野球におけるデータサイエンスの変化

野球では比較的古くからデータの活用はされており、スコアブックなどでイベントの記録などをし、それらを集計することで様々なことに応用をしてきた。James (1977) によっ

て野球におけるデータサイエンスが始まったとされるが、それにより何が変わったのかを説明する。例えば、長打率という指標を考えてみる。

$$\text{長打率} = \frac{\text{単打数} + 2 \times \text{二塁打数} + 3 \times \text{三塁打数} + 4 \times \text{本塁打数}}{\text{打数}}$$

これをみると、単純に塁を進める個数を重みとして計算している。しかし、二塁打は単打の2倍の価値があるのかと言えそこには異論があるかもしれない。その価値を数値化するためにデータサイエンスが必要となる。スポーツの分析においては絶対的な評価項目が作りやすいことが多く、野球では勝ちや得点などがそれに使われることが多い。例えば長打率で単純に二塁打は単打の2倍の価値としていたものが得点を基準にして、“得点価値”というセイバートリクス指標（データサイエンス指標）を用いると、ある年代の単打は“0.437”、二塁打は“0.786”と算出でき、実際には2倍よりも少し小さい価値であると解釈することができる（岡田ら、2017を参照）。このように従来記録されたデータを扱うことでデータサイエンスは十分に価値を見出してきた。

しかし、近年では測定技術の進化により、これまでは記録できなかったイベントなどを容易に記録できるようになった。まず、守備位置や打球の落下点の正確な位置（座標）が取得できるようになったのも技術の進化である。これにより、野球にはどんな影響があったのかと言え、それまではあまり注目されなかった（正確な評価が難しかった）守備の指標を作ることがこのデータ取得により可能となる。それはUZR（Ultimate Zone Rating）と呼ばれている。UZRは上述した得点価値も考慮して算出されるため、得点を基準とした守備の価値を評価する指標として使うことができる。現在はUZRをさらに改良した指標などの研究も進められている。

さらに近年の測定技術の進化で注目するのは打球や投球の回転軸や回転数までもが取得できるようになったことである。日本ではホークアイによりこれらの情報が取得が出来るようになった。日本の一部の球団の放送では投手の回転軸や回転数がテレビにも表示されるようになった。ただし、これらは数字を見てもそれが何を意味しているのか？という答えには辿り着きづらいかもしれないし、そこまで広く一般に浸透していないというのが現状である。データサイエンスにはそれら特徴を解釈可能となるような役割が求められており、それが実現すれば近い未来には投球・打球の回転軸や回転数を基にした戦術立案や効率的なトレーニング方法などにも活用されていくだろう。また、打球フォームや投球フォームなども画像や動画データとして保管することが可能となっている現代ではそれらを用いたトレーニングなどがすでに実施されている。とはいえ、得られているデータをどう扱えばという点ではまだ発展の余地は多くあり、ここにもデータを活用するデータサイエンスが広く求められている。

参考文献

- [1] James, B. (1977). *Baseball Abstract*.
- [2] Peter, N. (1974). *Concise survey of computer methods*, Petrocelli Books.
- [3] 岡田友輔ら (2017). 『プロ野球を統計学と客観分析で考える デルタ・ベースボール・レポート 1』, 水曜社.

卓球におけるデータサイエンスの活用と ゼロ変形交差一般化幾何分布による 打球回数のモデリング

中央大学 酒折 文武

1 卓球におけるデータ活用の現状

AIやIoT, 計測技術などの高度な発展に伴い, プロスポーツやオリンピックスポーツを中心としてデータ活用がますます進んでいる. 年間成績などのいわゆる“スタッツ”のみならず, プレイ毎に手動で記録されるログデータをはじめとして, センサーやカメラを用いてボールや選手の軌跡を収集したトラッキングデータ, そして非構造データである映像そのものなど, データの活用の幅は広がっている.

卓球の現場においては, 1 ラリー中の打球回数, ボールのバウンド位置, ボールの回転の種類などのラリーの情報をアナリストが映像から目視で入力し, その集計結果を選手の評価や戦術判断に用いることが一般的である. この中で, 打球回数の分布はサービスやラリーの強さや戦型を表す最も基本的な情報であり, Tリーグのウェブサイトでも選手毎の打球回数の分布のデータが公開されている. 打球回数に基づく研究としては例えば球回数と得失点に関する分析 (Tamaki *et al.*, 2013), 打球回数の分布に関する考察 (吉田他, 2015), 打球回数に対する幾何分布の当てはめと考察 (玉城・吉田, 2018), ゼロ変形一般化幾何分布 (Sastry *et al.*, 2016) の拡張を用いたモデリング (酒折・池袋, 2021) などがある.

近年ではビデオ映像からの深層学習による研究が徐々に広がりつつある. 例えば, 卓球台上部から撮影した映像に基づくボールの追跡と音によるバウンド位置の検出 (竹内・吉田, 2017), 卓球台側面から撮影した映像に基づく選手とボールの位置の検出 (Voeikov, *et al.*, 2020), プレイデータとトラッキングデータからの支配率や期待スコアなど算出・可視化 (Calmet *et al.*, 2023) などの研究がある. このような研究は, 前段落で述べたようなラリーの情報の自動抽出によるアナリストの業務軽減を目指しているものが多く, 姿勢解析などを用いたプレイの予測や選手評価などは今後の課題である.

本研究では酒折・池袋 (2021) で提案された二重ゼロ変形交差一般化幾何分布による打球回数のモデリングと, その応用について述べる.

2 打球回数のモデリング

卓球をはじめとしてテニス, バドミントンなどの多くのラケットスポーツでは, サービスを行う選手であるサーバーとそれをレシーブする選手であるレシーバーの間でのボールやシャトルの打ち合いにより得点を競う. このボールやシャトルの打ち合いをラリーという. 各ラリーにおけるサービスから得点までにボールを相手に打った回数を打球回数という.

打球回数はレットを除く正規のサービスと正規のリターンの和により表される。すなわち、打球回数が奇数の場合はサーバーの得点、偶数の場合はレシーバーの得点を意味する。

卓球では、前陣速攻型などのプレイスタイルをとる選手は打球回数が少なく、バックスピンドで相手に返してミスを誘うカット主戦型などのプレイスタイルをとる選手は逆に打球回数が多い傾向にある。このように打球回数の分布は選手やプレイの特徴を表すので現場でも重視されている。また、Tリーグのウェブサイトでは各選手ごとの打球回数の分布がグラフで表示されている。

ここでは、打球回数の分布として酒折・池袋 (2021) で提案されている二重ゼロ変形一般化幾何分布を当てはめる。この二重ゼロ変形交差一般化幾何分布 (double zero-distorted generalized cross geometric distribution, DZDGCG distribution) の確率関数は

$$f(x) = \begin{cases} 1 - p^{\alpha+1} & (x = 0) \\ p^{\alpha+1}(1 - q^{\beta+1}) & (x = 1) \\ p^{\alpha+\frac{x}{2}}q^{\beta+\frac{x}{2}}(1 - p) & (x = 2, 4, \dots) \\ p^{\alpha+\frac{x+1}{2}}q^{\beta+\frac{x-1}{2}}(1 - q) & (x = 3, 5, \dots) \end{cases}$$

で定義される。ここで $0 < p < 1$, $0 < q < 1$, $-1 < \alpha < \infty$, $-1 < \beta < \infty$ である。二重ゼロ変形交差一般化幾何分布 $DZDGCG(p, q, \alpha, \beta)$ では、卓球でいえば、サービスエースによりサーバー側の得点となる確率が $1 - p^{\alpha+1}$ 、レシーブエースによりレシーバー側の得点となる確率が $1 - q^{\beta+1}$ とし、それ以降はサーバー側の得点となる確率が打球ごとに $1 - p$ 、レシーバー側が $1 - q$ としたときの $X = \text{打球回数} - 1$ の分布を表していることになる。つまり、サービスエースのみならずレシーブエースにも特殊性をもたせていることになる。

酒折・池袋 (2021) ではこれらの確率分布の性質と、パラメータの最尤推定量、近似信頼区間を含む推測について述べている。例えば、 $DZDGCG(p, q, \alpha)$ からの独立な n 個の観測値 x_1, x_2, \dots, x_n に基づく p, q, α の最尤推定値は

$$\hat{p} = 1 - \frac{2n_a}{\sum_{x_i \geq 2} x_i - n_b}, \quad \hat{q} = 1 - \frac{2n_b}{\sum_{x_i \geq 2} x_i - 2n_a - n_b},$$

$$\hat{\alpha} = \frac{\log\left(1 - \frac{n_0}{n}\right)}{\log \hat{p}} - 1, \quad \hat{\beta} = \frac{\log\left(1 - \frac{n_1}{n - n_0}\right)}{\log \hat{q}} - 1$$

で与えられる。ここで、 $n_0 = |\{x_i | x_i = 0\}|$ (0 の個数), $n_1 = |\{x_i | x_i = 1\}|$ (1 の個数), $n_a = |\{x_j | x_j \text{は } 2 \text{ 以上の偶数}\}|$ (2 以上の偶数の個数), $n_b = |\{x_k | x_k \text{は } 3 \text{ 以上の奇数}\}|$ (3 以上の奇数の個数) である。

パラメータの信頼区間や AIC などについては酒折・池袋 (2021) を参照されたい。

3 打球回数の分布への当てはめ

Tリーグウェブサイトで公開されている 2021–2022 シーズン個人順位表 (男子シングルス) にリストアップされている選手の打球回数のデータを用いて、モデルのあてはまりの良さの評価と、選手間比較を行った結果の一部が図 1 である。ここで、各選手がサービス側のときのデータのみを用いている。左はランキング 1 位の張本智和選手 (右シェークドライ

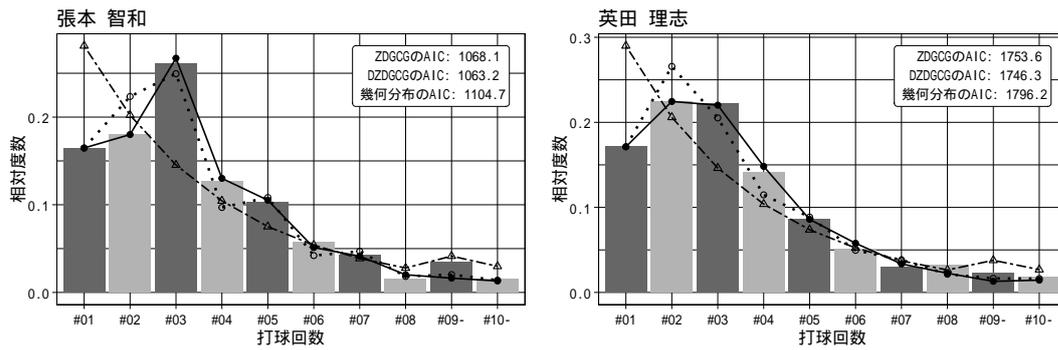


図 1: 2 選手の打球回数の分布

ブ型), 右はランキング9位タイの英田理志選手(右シェークカット型)の打球回数と当てはめた分布である. プレイスタイルが違い, 打球回数の分布の形状も異なるものの, いずれも黒い実線で表された二重ゼロ変形一般化幾何分布の当てはまりが非常に良いことがわかる. すなわち, 打球回数の特徴はほぼ二重ゼロ変形一般化幾何分布で表現できるといえる.

3.1 選手評価への応用

二重ゼロ変形一般化幾何分布のパラメータ p, q, α, β の推定値は各選手の特徴を反映している. とくに, $1-p$ は各打球での得点率を, q は相手の打球に対するラリー継続率を表している, すなわちそれぞれがオフェンス力, ディフェンス力に対応していると考えられる. また, α, β を $[0, 1]$ の範囲に制限した $\min\{\alpha + 1, 1\}, \max\{-\beta, 0\}$ はそれぞれサーブの得意さ, レシーブエースのされにくさに対応すると考えられる. これらのパラメータを用いて選手の評価を行うことが可能である.

3.2 選手のクラスタリングへの応用

打球回数の分布は選手により傾向が異なる. そこで, 当てはめた二重ゼロ変形一般化幾何分布の類似度を見ることにより, 打球回数の分布クラスタリングを行うことが可能である.

通常の変数データのクラスタリングとは異なりクラスタリングの対象は確率分布である. そこで, 分布間の距離としてKLダイバージェンスを用いることが考えられる. しかしながら, クラスタ分析における分布間距離としてKLダイバージェンスは非対称であり, かつ台集合の変換に対して不変であるなどの好ましくない性質を有している. そこで, 最適輸送に基づく Wasserstein 距離

$$W_p(f, g) = \left(\inf_{\gamma \in \Gamma(f, g)} \int \|x - y\|^p d\gamma(x, y) \right)^{1/p},$$

とくに $p = 2$ とした L_2 -Wasserstein 距離を用いる方法が考えられる.

ユークリッド距離に基づく k -平均法 (Lloyd のアルゴリズム) を Wasserstein 距離の場合に拡張したものとして Papayiannis *et al.* (2021) がある. しかしながら, Wasserstein

重心 (barycenter) は不適切な挙動をもち、かつ外れ値に対するロバスト性がとても低い。したがって、Lloyd のアルゴリズムのようにクラスター中心との距離に基づくのではなく、分布間 (対象間) の対距離に基づいた Wasserstein k -平均法 (Zhuang *et al.*, 2022) の方が適切であると考えられる。

詳細な分析結果や課題などについては当日述べる。

参考文献

- [1] Calmet, G., Erades, A. and Vuillemot, R. (2023). Exploring table tennis analytics: domination, expected score and shot diversity, *Machine Learning and Data Mining for Sports Analytics*, Sep 2023, Turin, Italy. hal-04240982.
- [2] Papayiannis, G.I., Domazakis, G.N., Drivaliaris, D., Koukoulas, S., Tsekrekos, A.E. and Yannacopoulos, A.N. (2021). On clustering uncertain and structured data with Wasserstein barycenters and a geodesic criterion for the number of clusters, *Journal of Statistical Computation and Simulation*, **91**(13), 2569–2594.
- [3] 酒折文武・池袋晴彦 (2021). 「ゼロ変形交差一般化幾何分布と卓球の打球回数への応用」『*計算機統計学*』 **34**(2), 59–75.
- [4] Sastry, D.V.S., Bhati, D., Rattihalli, R.N. and Gómez-Déniz, E. (2016). On zero-distorted generalized geometric distribution. *Communications in Statistics — Theory and Methods*, **45**, 5427–5442.
- [5] 竹内義則・吉田和人 (2017). 「体育館天井に設置された一台のカメラによる卓球のゲーム分析」『*スポーツ産業学研究*』 **27**(3), 265–275.
- [6] Tamaki, S., Yoshida, K. and Yamada, K. (2013). A short-time performance analysis in table tennis, *Int J. of Table Tennis Sciences*, **8**, 84–87.
- [7] 玉城将・吉田和人 (2018). 「卓球においてサービスがラリーに与える影響の定量化」『*名桜大学総合研究*』 **27**, 27–33.
- [8] Voeikov, R., Falaleev, N. and Baikulov, R. (2020). TTNNet: Real-time temporal and spatial video analysis of table tennis, arXiv:2004.09927.
- [9] 吉田和人・山田耕司・玉城将・内藤久士・加賀勝 (2015). 「卓球のワールドクラスの試合におけるラリーの特徴 – ラリー中の打球回数に着目して –」『*コーチング学研究*』 **28**(1), 65–74.
- [10] Zhuang, Y., Chen, X. and Yang, Y. (2022). Wasserstein K -means for clustering probability distributions, arXiv:2209.06975v2.

ヘルスデータサイエンスの現状と取り組み

横浜市立大学大学院データサイエンス研究科ヘルスデータサイエンス専攻 富田 誠

1 はじめに

2017年に滋賀大学で国内初となるデータサイエンス学部が開始され、翌2018年には首都圏にも横浜市立大学でデータサイエンス学部が開始された。以降、2019年に武蔵野大学データサイエンス学部、2021年に立正大学データサイエンス学部、2023年4月には一挙に5大学で、京都女子大学データサイエンス学部、順天堂大学健康データサイエンス学部、一橋大学ソーシャルデータサイエンス学部、名古屋市立大学データサイエンス学部、大阪成蹊大学データサイエンス学部が開始された(図1)。更に2024年には下関市立大学データサイエンス学部、宇都宮大学データサイエンス経営学部、千葉大学情報・データサイエンス学部などが予定されている。

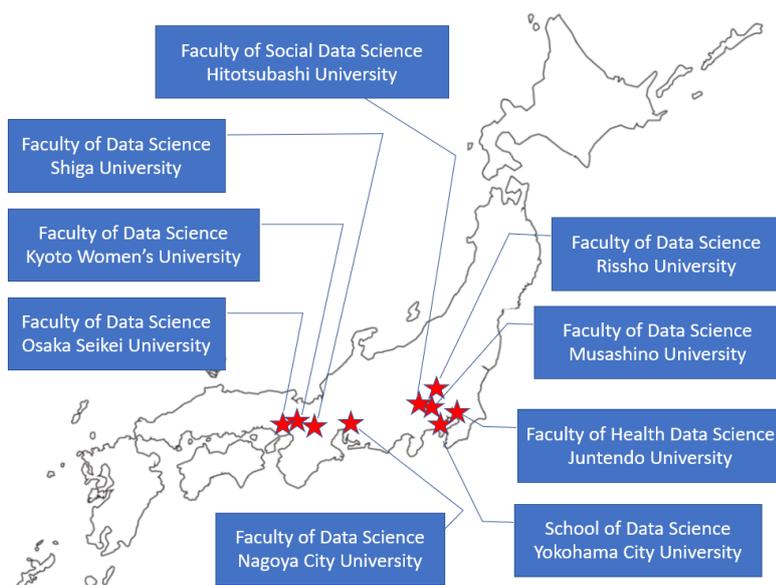


図 1: Data Science Faculties in Japan as of 2023 (update; Tomita, 2022)

一方、横浜市立大学には大学院データサイエンス研究科にヘルスデータサイエンス専攻が2020年に開始され、前述の通り、順天堂大学健康データサイエンス学部も有り、ヘルスデータサイエンスに取り組む学部や大学院研究科が出現し始めた。横浜市立大学は2022年までは医学部とデータサイエンス学部が有る唯一の大学であったが、前述の名古屋市立大学も両学部を有する。本専攻の状況や取り組みを紹介し、特色や特徴を考察したい。

2 ヘルスデータサイエンス専攻の紹介

2.1 3つのコア・スキル

横浜市立大学には大学院データサイエンス研究科ヘルスデータサイエンス専攻には重要な柱として「3つのコア・スキル」が設定されており、ヘルス領域の専門知識（いわゆるドメイン知識）、研究デザイン学、そして著者も科目担当している生物統計学である。これら3つがバランス良く習熟されることを目指しており、本専攻の基礎教育といえる。

2.2 5つの重点分野

続けて同専攻では特にフォーカスする研究分野として、Real World Evidence Research, Health Economics & Outcomes Research, Evidence-Based Health Administration, Design & Analysis of Clinical Trial, Primary Care Research と5つの重点分野が有り、特に横浜市の状況に合う領域も挙げられている。

2.3 Opportunity

主に、医療機関、行政機関、ヘルス関連企業（製薬企業、薬品会社、CROなど）、ヘルスケア系シンクタンク、他で挙がってくるようなリサーチクエスチョンを解決していくようなヘルスデータサイエンス・リサーチに取り組んでいる。

3 ヘルスデータサイエンス研究

当日は現状や取り組みなどについて紹介する予定である。

参考文献

- [1] Tomita, M. (2022) Health Data Science Research and Education Initiatives in Japan. European Conference of Data Analysis 2022, Naples, Italy.
- [2] 栗原考次, 富田 誠, 和泉志津恵, 岩崎 学 (2023) データサイエンス学部：最近の状況と新時代への展開, 日本分類学会第42回大会予稿集, 39-40.

医療リアルワールドデータを用いた薬剤耐性対策における 臨床疫学・薬剤疫学研究と今後の薬学教育への展開

京都薬科大学 臨床薬剤疫学分野 村木 優一

1. はじめに

抗菌薬耐性菌の増加は、公衆衛生上の世界的な問題となっており、世界各国で国家行動計画が策定されている。我が国においても2016年に薬剤耐性（AMR）対策アクションプランが公表され、2023年に改訂された。6つの目標を掲げるAMR対策アクションプランの1つに「動向調査・監視」がある。これまで、日本では抗菌薬の使用状況の把握はなされていなかったが、我々は販売量およびレセプト情報・特定健診等情報データベース（NDB）やDPC（診療群分類包括評価）情報を用いて医療機関や地域におけるサーベイランス体制の構築に貢献し、今後のAMR対策を進める上で大きな成果を果たしている。

AMR対策を実行し、評価する上で重要なことは、耐性菌や抗菌薬の使用状況を地域間や医療機関毎に経年的に調査することが重要である。また、種々の対象集団における薬物治療の効果や副作用を評価し、医療の質を高めることも必要である。近年、こうした課題に対して医療リアルワールドデータの利活用が求められているが、使用する情報源に応じて利点欠点があることを理解している研究者や政策立案者はほとんどいない。しかしながら、我々はこれまでに様々な情報源を用いて、種々の目的に応じた抗菌薬の使用状況を明らかにし、方法論の確立や継続した調査の実行に関わる根拠を創出してきたため、紹介する。

2. 様々な情報源を用いた日本における抗菌薬使用動向の把握と調査体制の構築

抗微生物薬だけでなく、すべての医薬品における使用状況を把握するには、尺度を揃える必要がある。医薬品はそれぞれ、決められた投与量や投与方法が決められている。また、投与期間も決められており、特に抗微生物薬は感染症毎に推奨される投与期間が異なる。さらに、医薬品には先発医薬品や後発医薬品が存在するように複数の製薬会社から販売され、1つの医薬品でさえ、錠剤、バイアルに含まれている含有量が異なる規格が複数存在する。こうした複雑に異なった条件下において同じ尺度で成分や系統別に表現しなければならない。

医療リアルワールドデータの情報源には、様々な種類がある。それらの構造化データに医薬品の情報も含まれるが、必ずしも分析に適した状態ではない。そのため、目的に応じて分析に必要な情報を追加しなければならないが、簡便な作業ではない。また、これらの情報源は迅速性や網羅性など利点欠点が存在する。こうした問題を把握した上で我々は、まず医療機関における電子カルテ情報を用いて注射用抗菌薬の使用動向を調査した¹。その後、販売量情報²、薬局情報³、NDBを用いて経年的な日本全国⁴、都道府県⁵、二次医療圏^{6,7}における抗菌薬の使用状況を明らかにした。また、これらの経験を活かし、システム化することで各医療機関や日本における抗菌薬の経年的な使用動向の把握が可能となり、AMR対策に役立てられている。

3. 大規模保険請求情報を用いた薬物治療および保険医療制度の評価

日本は国民皆保険制度であり、施設機能や診療行為に対して報酬が定められ、医療が行われている。一方、日々、医療は進歩し、様々な知見に基づき治療内容は変化している。一般的に臨床研究では、対象期間における集団に対して調査対象を抽出して評価されるが、それぞれの患者における施設背景や診療を受けた時期が考慮されることはほとんどない。しかしながら、医療リアルワールドデータの1つである大規模保険請求情報は、長期間における種々の施設背景で治療や保険請求を受けた患者に対する分析が可能である。そのため、我々は薬物治療において必要とされる診療行為に紐づく診療報酬を利用し、効果や副作用への医療者の貢献を評価している⁸⁻¹²。

バンコマイシン⁹やテイコプラニン¹⁰、アルベカシン¹¹、ポリコナゾール¹²という抗微生物薬は、治療有効濃度と副作用発現濃度の間隔が狭く、患者の状態や患者間で体内における薬物動態が大きく異なる。そのため、これらの薬物を投与する場合には、適切な時期に採血し、血液中の薬物濃度を測定し、適切な投与計画を解析した上で投与量や投与方法を管理する必要がある。また、これらの投与計画に関わるのは主に薬剤師であり、介入した場合に特定薬剤治療管理料といった診療報酬が請求できる。一方、薬剤師は個々の患者における薬物治療の管理・指導を行った場合に薬剤管理指導料、病棟に薬剤師が配置されている場合に病棟薬剤業務実施加算が請求できる。

我々は、これらの薬物に対して算定された各診療報酬の請求状況の経年変化を追うことで、薬剤師による関与状況を視える化した。また、診療報酬の請求状況により算定できていない施設背景や治療効果や副作用に与える要因を評価した。このような評価方法は、様々な領域で応用可能であり、今後、多方面における診療報酬制度の評価や、医療者の貢献や過不足を視える化し、医療の質向上に活用することが期待できる。

4. さいごに

2024年度の医学部、歯学部、薬学部の入学生に対して改訂された教育モデル・コア・カリキュラムが適応される。そのなかの項目には、データサイエンスやビッグデータ、大規模保険請求情報の利活用といった項目が含まれ、これからの医療系大学では、未来の社会や地域を見据え、多様な場や人をつなぎ活躍できる医療人を養成できる教育体制を構築することが求められている。一方、現状の医療系大学において学生に教育できる人材が不足しているのが現状である。また、医薬品に関する情報は、医薬品がどのように患者に対して使用されるか理解できることが不可欠である。

医療系大学の学生は、実務実習を行うものの、実際の医療現場で活躍するのは、国家試験の取得後であり、低学年の学生に医療を理解させる環境に乏しいのが現状である。また、他学部で医療リアルワールドデータを取り扱う場合には、医療そのものを理解する機会がほとんどない。こうした課題を解決する1つの手段に医療リアルワールドデータがあると考えます。我々は、現在、京都女子大学とも連携し、医療リアルワールドデータを用いた薬学部における教育研究

体制を構築し、運用を開始している。生成系 AI や DX の技術が急速に進む今日において、質の高い教育研究を展開するには、大学間の枠を超え、より多くの医療従事者と共に医療リアルワールドデータを用いた教育研究を推進すべきである。

参考文献

- [1] Yuichi Muraki, Masaki Kitamura, Yorinobu Maeda, Takashi Kitahara, Takeshi Mori, Hidehumi Ikeue, Masami Tsugita, Koji Tadano, Katsutoshi Takada, Takashi Akamatsu, Takehiro Yamada, Tadashi Shiraishi, Masahiro Okuda. Nationwide surveillance of antimicrobial consumption and resistance to *Pseudomonas aeruginosa* isolates at 203 Japanese hospitals in 2010. *Infection*, 41: 415-423, **2013**.
- [2] Yuichi Muraki, Testuya Yagi, Yasuhiro Tsuji, Nobuhiro Nishimura, Masaki Tanabe, Takashi Niwa, Tamayo Watanabe, Shuhei Fujimoto, Kazuro Takayama, Nobuo Murakami, Masahiro Okuda. Japanese antimicrobial consumption surveillance: first report on oral and parenteral antimicrobial consumption in Japan (2009–2013). *J Glob Antimicrob Resist*, **2016**.
- [3] Yuichi Muraki, Masayuki Maeda, Ryo Inose, Koki Yoshimura, Naoki Onizuka, Masao Takahashi, Eiji Kawakami, Yoshiaki Shikamura, Naotaka Son, Makoto Iwashita, Manabu Suzuki, Masayuki Yokoi, Hirokazu Horikoshi, Yasuaki Aoki, Michiyo Kawana, Miwako Kamei, Hajime Hashiba, Choichiro Miyazaki. Exploration of Trends in Antimicrobial Use and Their Determinants Based on Dispensing Information Collected from Pharmacies throughout Japan: A First Report. *antibiotics*. 11(5), **2022**.
- [4] Daisuke Yamasaki, Masaki Tanabe, Yuichi Muraki, Genta Kato, Norio Ohmagari, Tetsuya Yagi: The First Report of Japanese Antimicrobial Use Measured by National Database Based on Health Insurance Claims Data (2011-2013): Comparison with Sales Data, and Trend Analysis Stratified by Antimicrobial Category and Age Group, *Infection*. 46: 207-214, **2018**.
- [5] Yoshiki Kusama, Masahiro Ishikane, Chika Tanaka, Yuki Kimura, Erina Yumura, Kayoko Hayakawa, Yuichi Muraki, Daisuke Yamasaki, Masaki Tanabe, Norio Ohmagari. Regional variation of antimicrobial use in Japan from 2013-2016, as estimated by the sales data. *Jpn. J. Infect. Dis.* 72(5), 326-329, **2019**.
- [6] Koizumi Ryuji.; Kusama Yoshiki.; Muraki Yuichi.; Ishikane Masahiro.; Yamasaki Daisuke.; Tanabe Masaki.; Ohmagari Norio. Effect of population inflow and outflow between rural and urban areas on regional antimicrobial use surveillance. *PLOS ONE*. 16(3):e0248338, **2021**.
- [7] Kanako Mizuno, Ryo Inose, Yuna Matsui, Mai Takata, Daisuke Yamasaki, Yoshiki Kusama, Ryuji Koizumi, Masahiro Ishikane, Masaki Tanabe, Hiroki Ohge, Norio Ohmagari, Yuichi Muraki. Search for Indexes to Evaluate Trends in Antibiotic Use in the Sub-Prefectural Regions Using the National Database of Health Insurance Claims and Specific Health Checkups of Japan. *Antibiotics*. 11(6), **2022**.
- [8] Yuichi Muraki, Yoshiki Kusama, Masaki Tanabe, Kayoko Hayakawa, Yoshiaki Gu, Masahiro Ishikane, Daisuke Yamasaki, Tetsuya Yagi, Norio Ohmagari. Impact of antimicrobial stewardship fee on prescribing for Japanese pediatric patients with upper respiratory infections. *BMC Health Services Research*. 20(1), **2020**.
- [9] Ryota Goto, Yuichi Muraki, Ryo Inose, Yoshiki Kusama, Akane Ono, Ryuji Koizumi, Masahiro Ishikane, Norio Ohmagari. Influence of pharmacists and infection control teams or antimicrobial stewardship teams on the safety and efficacy of vancomycin: A Japanese administrative claims database study. *PLOS ONE*, 17(9), **2022**.
- [10] Ryota Goto, Yuichi Muraki, Ryo Inose, MoenoIchii, Keisuke Sawada, Kanako Mizuno, Ryuji Koizumi, Shinya Tsuzuki, Masahiro Ishikane, Norio Ohmagari. Trends in teicoplanin loading dose implementation from 2010 to 2019 and evaluation of safety and efficacy factors: a retrospective cohort study based on a Japanese administrative claims database. *Journal of Pharmaceutical Health Care and Sciences*, **2023**, *in press*

- [11] Saki Ito, Ryota Goto, Ryo Inose, Yoshiki Kusama, Akane Ono, Ryuji Koizumi, Masahiro Ishikane, Norio Ohmagari, Yuichi Muraki. A study of trends and factors associated with therapeutic drug monitoring(TDM) implementation for arbekacin treatment using a large Japanese medical claims database. *J. Infect. Chemother.* 28(9), 1266-1272, **2022**.
- [12] Ryo Inose, Ryota Goto, Misato Shimoki, Norio Ohmagari, Yuichi Muraki. Trends in implementation of loading dose of voriconazole from 2010 to 2019 and influencing factors of loading dose based on a Japanese administrative claim database: A retrospective cohort study. *Biol. Pharm. Bull.* **2023**, *in press*

臨床試験を哲学する

京都府立医科大学大学院医学研究科 生物統計学 手良向 聡

統計学は積分や行列など数学で用いる道具を用いるが、数学の延長線上にはない学問である。ただし、数学の一分野である確率論は、1500年代にジロラモ・カルダーノ、1600年代にブレイズ・パスカルとピエール・ド・フェルマーが議論した賭け金の分配問題（キース・デブリン「世界を変えた手紙」）に端を発しており、そこからヤコブ・ベルヌーイによる大数の法則、アブラム・ド・モアブルによる正規分布などが生まれた。したがって、確率論を土台にして数理統計学という学問が生まれたといえる。統計学は確率を扱う学問であるが、確率とは何かについては多くの考え方がある。また、統計学は広く社会で利用されているが、各領域における知識がないと結果の正しい解釈ができない。とりわけ、意思決定を前提として用いることが多いことから、人間の認識と深く係わるのが特徴的である。

医学は普遍的な真理を追求する科学の一分野であるが、医療は個人に対して最適な技術を選択して適用することが要求される。医学から生み出された先端技術は、主に統計学に基づく科学的方法を用いて相対的に評価され、医療に適用される。統計的な技術評価を医学または医療と関連付ける際には、2つの大きなギャップを認識する必要がある。医学と技術評価の間のギャップは、「決定論」対「非決定論」である。技術評価と医療の間のギャップは、意思決定の場に関係し、「対集団の確率」対「対個人の確率」である。後者のギャップは、確率とは何かという問いと直接的に関係する（ドナルド・ギリース「確率の哲学理論」など）。確率の意味論ではなく、「不確かさ」を分類するという観点もある（Tony O'Hagan, Significance 2004:132-133）。そこでは、事象に対する世界ベースの確率を「偶然的（aleatory）不確かさ」と呼び、命題に対する情報ベースの確率を「認識的（epistemic）不確かさ」と呼ぶ。「統計学は、世界のある量について私たちが認識的不確かさを有しているときに用いられ、その未知の量をパラメータと呼ぶ（David Spiegelhalter「The Art of Statistics」）」という説明は理にかなっていると考える。また、哲学の立場からは「統計学は、一定の存在論的前提の上に立つ科学的認識論である」と説明されている（大塚淳「統計学を哲学する」）。なお、医学と統計学の両方に関係の深い思想として、プラグマティズムが挙げられる（伊藤邦武「パースのプラグマティズム」）。

臨床試験の方法論は、ロナルド・フィッシャーが考案した統計的実験（技術的実験）の方法論を基礎として発展してきた。人工的に作った純粋な条件下での因果関係を立証しようとする科学的実験に対して、統計的実験は以下の特徴を有する（竹内啓「統計学的な考え方－デザイン・推測・意思決定－」）。

- ・ 実験の場は、現実の応用の場に近い状況に設定される。
- ・ 結果の分析には、誤差の存在を前提にしなければならない。
- ・ いくつかの因子を同時に変化させて結果を見る必要があることがある。

- ・ 目的は、何らかの基準によって現実の場において最も良い結果が得られるような条件を求めることである。

なお、結果を再現性によって保証することが可能な基礎医学・生物学研究（科学的実験）と異なり、臨床試験（統計的実験）は、同じ条件で繰り返すことが困難であり、デザインと手続きの正しさをしか結果を保証できない。

ベイズ流統計学は、古典的と呼ばれている頻度流統計学に対して異端の統計学として扱われてきた歴史をもつ（シャロン・バーチュ・マグレイン「異端の統計学ベイズ」）。米国FDA（Food and Drug Administration）は、2010年に「医療機器の臨床試験におけるベイズ流統計学の利用に関するガイダンス」を公表し（邦訳：手良向聡「なぜベイズを使わないのか!？」）、その後、臨床試験にベイズ流接近法が有用であるという報告が増えてきている。ベイズ流統計学を臨床試験に適用する際には、臨床家と統計家の真の協同が不可欠である。臨床家は、事前情報の確かさ、目標値の見積もり、実施可能性などについて、専門家としての見解をもつ。統計家は、事前情報、目標値、確率閾値の妥当性などについて臨床家と議論し、標本サイズを含む試験デザインを決定しなければならない（文献 [1]）。ちなみに、臨床試験に責任をもつ統計家は試験統計家と呼ばれ、日本計量生物学会は2017年から試験統計家認定制度を設立・運用している。

本講演のまとめを以下に述べる。

- ・ 臨床試験は試験計画の重要性が特徴的であり、統計的実験に特有の方法論を用いるべきである。
- ・ 技術評価という観点から、その方法論はプラグマティズム（主に可謬主義）、およびベイズ流認識論につながる。
- ・ 信念に関する確率の計算についてはベイズ流統計学、意思決定を行いたい場面では頻度流統計学が有用である。
- ・ 実践においては、ベイズ流統計学は頻度流統計学を包含するという見方もあると考える。

参考文献

- [1] 手良向聡 (2023). ベイズ流臨床試験における標本サイズ設定：2つの事前分布を用いた推論上の性能に基づく接近法. 計量生物学, 44, 35–51.

長崎大学 情報データ科学部のあゆみとこれから

長崎大学 情報データ科学部長 西井 龍映

1 はじめに

2020年4月に創設された長崎大学情報データ科学部は、それまでの工学部6コースのうち情報工学コースを切り出し、あらたにデータサイエンス系の教員群を受け入れて組織されました。データサイエンス系の教員は教育学部・ICT基盤センターや他大学・理化学研究所および民間企業から採用しました。学生定員は情報工学コースの定員50名に加えて、教育学部から60名を移管した110名でした。2023年度からは経済学部夜間主コースの募集停止に伴い、10名増となりました。一学部、一学科、2コース（データサイエンスコース、インフォメーションサイエンスコース）で運営しています。教員数は30名で、現在2名を公募中です。2023年は学部創設4年目に当たり、一期生が卒業します。情報データ科学部のこれまでのあゆみとこれからについて、述べさせていただきます。

2 新学部の創設準備 (2018～2019)

筆者は2018年5月当時は九州大学マスコアインダストリ研究所に在職していましたが、「長崎大学に情報系新学部の創設」に協力依頼があり、引き受けることにいたしました。当時決定すべき最も重要な事項は、学部のディプロマポリシーを理系型か文理融合型かを定めることでした。新学部のカリキュラムはおおよそ情報工学コースのそれを引き継ぐことが決まっていました。ただ情報工学コースのカリキュラムでは文系型入試で入学してきた学生が苦勞するだろうと考えました。また当時すでに同系学部を創設済みだった滋賀大学、広島大学、横浜市立大学を訪問し、実情をヒアリングしてきました。その結果からも理系学部にすることになりました。その後の社会的要請の変化もあり、情報データ科学部は2024年度入試から文系型も導入することになりました。

3 学部創設とコロナ禍 (2020～2022)

学部創設はまさにコロナ禍が猛威を振るいかけているときでした。新入生には10日間程度の入学ガイダンスでオンライン授業の受講方法を周知したのち、登学禁止となりオンライン授業となりました。新設学部には先輩もいません。特に留学生や長県外からの入学者は、長崎での新しい人間関係を構築することが難しいと予想されました。そこで4人程度のチューターグループで友人関係を作ってくれることを期待しました。また新たに赴任した10名の教員には、オンライン教育システムに慣れていただくことに注力しました。

長崎大学は当時の河野学長をはじめ感染症の専門家が多く在籍しています。大学執行部はコロナ禍は長期にわたること、また周期があることを見越し、独自のコロナ対策を行ってきました。この対応は他大学に誇れるものです。

コロナ禍で再認識したことは、人と自由に面談することや自由に旅行すること、すなわち人と人との繋がりのおかげです。またオンライン講義・会議システムの導入が一挙に進みましたし、学会・シンポジウムがオンライン会議に移行したことででしょうか。ちなみに2021年の統計関連学会連合大会は長崎大での開催予定でしたが、オンライン開催となりました。

4 実課題解決プロジェクト（PBL）（2020～）

「探求学習」は2020年から小学校、2021年から中学校、2022年からは高等学校でスタートしています。情報データ科学部は探究学習の大学版であるPBLを、学部1,2年生の必修、3,4年生の選択科目として設定しました。コロナ禍で企業担当者と学生の対面授業が制約されるなどしましたが、現在は30社以上の企業・自治体と課題発見・解決にむけてグループ学習を続けています。PBLの当初の目的の一つに、課題提供会社へ就職するという4年間にわたる長期インターンシップとの捉え方がありました。ただPBLで生きた課題に接することで、社会で働くことの意義を発見することに繋がったようです。課題を提供した企業に就職するという以上の職業意識を高める意味がありました。

5 国際交流（2022～）

本学部は国際化を目指し留学生の入学定員10名を設け、外国人教員を3名採用しています。ただコロナ禍のため2名の外国人教員が赴任辞退となりました。2022年からは泰日工業大学（タイ）に国際PBLで学生14名、10名（2023年）を派遣しました。また（2022年にはさくらサイエンスで教員・学生8名をインドから招聘し、VIT（Vellore 工科大学）、Birla 理工大学との交流協定を結びました。また同年から「インドとの交流による情報データ科学部の国際展開力強化事業」を開始し、2023年にVITから9名の代表団を招聘し、12月から7名の学生が4-6週間滞在して、インターンシップを行います。

6 おわりに

情報データ科学部は来年度から下記のようなシステム変更を行います。

- 情報データ科学部を基礎学部の一つとする総合生産科学研究科が設置、2024/4
- 長崎スタジアムシティのサテライトキャンパスに7研究室が移転、2024/10
- インドの2大学と教員・学生の相互派遣を開始、2024/4
- 推薦入試に10名の女性枠を設置、2025/4
- 経済学部との共修プログラムを開始、2025/4

これからも長崎大学情報データ科学部へのご支援をよろしくお願いいたします。

データサイエンスにおける時空間統計科学の役割

東京大学名誉教授 矢島 美寛

1 序

データサイエンスにおける時空間統計科学の役割について講演する。ただし以下の内容は最近この問題に対して抱いている所感であり、確固とした見解にはまだ至っていない。ご出席の方々から忌憚のないご意見をいただければ幸いである。

また時空間統計科学がカバーする理論・応用の範囲は膨大であり、本講演の内容は網羅的あるいは中立的ではなく、選択的である。あくまで発表者の能力で叙述できる範囲および興味範囲に限られていることをあらかじめ断っておく。

2 データサイエンスとは何か？

いくつかの文献およびデータサイエンス関係の学部が設置されている大学のホームページを拝読・拝見した。それらを参考に発表者なりに「データサイエンス」を定義すれば、「情報科学・統計科学を分析の基盤として、客観的データ（証拠）に基づき、学問各分野（人文・社会科学分野から自然科学分野にいたるまで）が抱える課題の解決を図り、新たな知見を得る科学」となる。特に「ビッグデータの時代」といわれる今日、いわゆる「文理」の垣根を越えて「融合」したサイエンスとしてその重要性はいやが上にも増している。次節では統計科学の一分野であり、21世紀に入り大きな発展を遂げている時空間統計科学について概説する。次々節ではそれがデータサイエンスにおいて果たすべき役割、今後の課題について述べる。

3 時空間統計科学

近年新たな発展を遂げている統計科学諸分野の中でも、時空間統計科学は最も注目を浴びている分野の1つである。その理由としては環境学・疫学・経済学・地域科学を含む広範な学問分野においてグローバルな視点からその時空間的変動メカニズムを解明しなければならぬ喫緊の問題に我々が直面していることにある。

ここで時空間データの数学的表現およびデータを解析するモデルについて、簡単に説明する。実数の全体を $\mathbf{R} = (-\infty, \infty)$ とし、その d 次元 Euclid 空間を \mathbf{R}^d と表す。また整数の全体を $\mathbf{Z} = \{0, \pm 1, \pm 2, \dots\}$ とし、その d 次元直積集合 $\mathbf{Z} \times \dots \times \mathbf{Z} = \{(z_1, z_2, \dots, z_d) | z_i \in \mathbf{Z}, i = 1, 2, \dots, d\}$ を \mathbf{Z}^d と表す。 \mathbf{R}^d と \mathbf{Z}^d を統一的に表す場合には \mathbf{K}^d とする。次に観測地点・時点 (site) を $\mathbf{s} = (s_1, s_2, \dots, s_d)' \in \mathbf{K}^d$ とする。なお簡単のため、時点を含めて「地点」とよぶ場合もある。

そのとき $Y(\mathbf{s})$ を確率変数あるいは確率ベクトルとし, \mathbf{s} において観測されるデータとする. たとえば $d=2$ のときは, s_1 を緯度, s_2 を経度として, $Y(\mathbf{s})$ はその地点における地価などを考えればよい. $d=3$ のときは, s_3 を高度として, その地点の気温などとする. 時空間データの場合は $d=4$ とし, s_4 を時点とする. なお $d=1$, s_1 を時点とすれば時系列データである. \mathbf{s} の動く範囲を D としたとき, 確率変数あるいはベクトルの集合 $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ を確率場 (random field) という.

以上の数学的表現に基づいて, 時空間データは, 地点参照データ (point reference data), 地域 (格子) データ (areal(lattice) data), 点配置データ (point pattern data) の3種類に大別される. まず地点参照データは, D が正の面積をもつ d 次元直方体を含む \mathbf{R}^d の部分集合であり, \mathbf{s} は D 上で連続的に変化する. 先ほどの気温や風速・風向などがこのカテゴリーに属する. 次に格子データあるいは地域データでは, D が高々可算個の地点からなる. 地価あるいは各格子状に画素が与えられた画像データなどの例がある. 最後に点配置データでは観測地点が確率変数になる. たとえば地震などの事象が生じた地点を解析する場合である. \mathbf{s} で事象が生じたとき $Y(\mathbf{s}) = 1$, 生じなかったとき $Y(\mathbf{s}) = 0$ とする. このとき $\mathbf{N} = \{\mathbf{s} : Y(\mathbf{s}) = 1, \mathbf{s} \in D\}$ が生じた地点の全体である. 地震であれば震度を $M(\mathbf{s})$ とし, \mathbf{N} と $\{M(\mathbf{s}) | \mathbf{s} \in \mathbf{N}\}$ の関連などについて解析する. $M(\mathbf{s})$ を \mathbf{s} におけるマーク (mark) という.

本講演では地点参照データに対するモデルを例として取りあげ, その現状とデータサイエンスにおける今後の役割について展望する. 取りあげるモデルは時系列モデルの一般化である定常確率場, 非定常確率場としては固有定常確率場である. 格子データおよび点配置データに対するモデルは割愛するが興味のある方は発表時の参考文献を参照されたい.

4 データサイエンスにおける時空間統計科学

思いつくままに列挙する.

(1) 「ビッグ・データ」の時代において, $Y(\mathbf{s})$ を関数とする Spatial Temporal Functional Data Analysis (STFDA) は理論だけではなく, 実際のデータ解析においても feasible となり, 重要性を増していると思う. 「究極」の高次元 (無限次元) データ解析である.

(2) 現在までの時空間統計モデルは1変量データに対するモデルが多い. 多種多様なデータが入手出来る今日, 多変量時空間モデルの構築とその応用可能性の解明が必要となる.

(3) Gaussian processes は時空間モデルの "building block" であるが, より "realistic" な Non Gaussian processes の開発が必要となるデータが存在する. 例えば降雨量データなどがある (cyclonic patterns, squall lines).

(4) 機械学習・深層学習で応用されているニューラルネットワークとは関係があるのかないのかの解明が必要である. たとえば(1)に関連して, 時空間データが「入力」の場合, 「出力」はどのような関数となり, どのような時空間相関構造をもつのであろうか?

(5) 最後にビッグデータに対して, 統計的漸近理論はどのような意義をもつであろうか? データが大量にある世界では, 漸近理論がより現実的な貢献をなすということであろうか? ただし漸近理論の有効性は何らかの「定常性」の前提のもとで成立する. 定常性の成立しない世界では, 漸近理論はもはや現実的な意味をもたない. 長年漸近理論を研究してきた発表者としては前者に期待をかけたい.

データサイエンスと医学統計学 ～統計的因果推論の視点から～

統計数理研究所・順天堂大学 岩崎 学

1 はじめに

データサイエンスは、社会のあらゆる分野において、データに基づく価値創造に寄与するが、中でも医学分野では、Evidence-Based Medicine (EBM) の言葉が表すように、データという客観的な存在を通じて医療あるいはさらに広く健康全般に貢献している。医学分野では（あるいは他の多くの分野でも）ある処置の効果の立証は重要な研究課題である。降圧剤は狭心症や心筋梗塞などの心疾患の予防に有効か？マスクの着用は新型コロナの蔓延を防ぐか？など様々な例がある。医学分野以外でも、授業への出席は試験の点数の上昇に寄与するか？といった身近な例もある。

以下では、降圧剤の服用、マスクの着用、授業への出席などの原因系を一般的な用語として処置 (treatment) といい、結果系（心疾患予防、コロナの蔓延予防、試験の点数の上昇）を結果あるいは結果変数 (outcome) という。

2 統計的因果推論

処置効果の立証では、比較が重要であるし、比較なしには処置効果の立証はできないと言っても過言ではない。処置効果立証のための統計学上の方法論が統計的因果推論 (statistical causal inference) である。因果推論にはいくつかのアプローチの仕方があるが、ここでは潜在的アウトカム (potential outcomes) に基づく議論を紹介する。統計的因果推論について詳しくは阿部・岩崎 (2022)、岩崎 (2015) あるいはそれらの巻末の参考文献を見られたい。

ここでは簡単のため、処置および結果変数はともに2値とし、処置 (treatment) あるいは対照 (control) を表す変数を Z ($Z = 1$: 処置, $Z = 0$: 対照) とし、結果変数を表す変数を Y ($Y = 1$: 効果あり, $Y = 0$: 効果なし) とする。 Z が Y に及ぼす影響 ($Z \rightarrow Y$) が推定対象である。個体 i に対し、処置 ($Z_i = 1$) のときの潜在的な結果 (potential outcomes) を $Y_i(Z_i = 1)$ (あるいは単に $Y_i(1)$)、対照のときの潜在的な結果を $Y_i(Z_i = 0)$ (あるいは単に $Y_i(0)$) とする。処置の下では $Y_i(1)$ が観測され、対照の下では $Y_i(0)$ が観測されるが、それらは同時には観測されない。個体 i は、潜在的アウトカムの組 $\{Y_i(0), Y_i(1)\}$ で特徴付けられる。このとき、個体処置効果 (individual treatment effect) は $\tau_i = Y_i(1) - Y_i(0)$ と定義されるが、これは観測も推定もできない量である。個体処置効果の母集団全体での期待値を $\tau = E[Y(1) - Y(0)]$ と定義し (definition), これを平均処置効果 (average treatment effect : ATE) と言い、これが我々の推定対象 (estimand) である。

ATE τ が推定できるための条件 (identifiability), および具体的な推定法 (estimation) は、上述の書物などを参照されたい。

3 因果媒介分析

処置効果の立証だけでは研究課題に答えきれていない場合が多い。何故降圧剤は心疾患予防に有用なのか、何故マスク着用はコロナの蔓延予防になるか、何故授業に出席すると試験の点数が高いのか、といった疑問に答える必要がある。そのための方法論が因果媒介分析 (causal mediation analysis) である。記号的には、処置 (Z) から結果 (Y) への因果経路 (causal path) の中間に媒介因子 (mediator) M を考え、 $Z \rightarrow Y$ に加え、 $Z \rightarrow M \rightarrow Y$ の経路を想定する。すなわち、降圧剤 (Z) \rightarrow 血圧下降 (M) \rightarrow 心疾患予防 (Y)、マスク着用 (Z) \rightarrow 飛沫防止 (M) \rightarrow コロナ予防 (Y)、出席 (Z) \rightarrow 課題 (M) \rightarrow 試験 (Y) の経路を想定し、因果のメカニズムを探ろうとする。そして、 M を想定しない $Z \rightarrow Y$ の効果を総合効果 (total effect)、 $Z \rightarrow M \rightarrow Y$ の効果を間接効果 (indirect effect) あるいは媒介効果 (mediation effect) という。また、 M を想定するがそれを介さない効果を直接効果 (direct effect) という。

潜在的アウトカムを用いた媒介分析は次のように定式化される。まず、新たな記号として、処置が $Z = z$ 、媒介因子が $M = m$ のときの結果変数を $Y(z, m)$ とする。個体 i に対し、処置 Z_i の下での媒介因子を $M_i(Z_i)$ とすると、結果変数は $Y_i(Z_i, M_i(Z_i))$ と表される。ここで、 $Y_i(1) = Y_i(1, M_i(1))$ 、 $Y_i(0) = Y_i(0, M_i(0))$ であり、総合効果 (total effect) は

$$TE = E[Y_i(1) - Y_i(0)] = E[Y_i(1, M_i(1)) - Y_i(0, M_i(0))] \quad (1)$$

となる。ここで、 $Y_i(1, M_i(0))$ を考える (ここが重要)。そして、

$$NIE = E[Y_i(1, M_i(1)) - Y_i(1, M_i(0))] \quad (2)$$

$$NDE = E[Y_i(1, M_i(0)) - Y_i(0, M_i(0))] \quad (3)$$

と定義し、NIE を自然間接効果 (natural indirect effect)、NDE を自然直接効果 (natural direct effect) という。(1), (2), (3) より $TE = NIE + NDE$ の等式が成り立つ。「自然」(natural) とは、処置が与えられるとそれに応じて自然に媒介因子の値が定まる、すなわち媒介因子に介入して $M_i(Z_i) = m$ と強制的に設定したのではないという意味である。

因果媒介分析の文脈では、因果推論のときと同様、効果の定義 (definition)、識別性 (identification)、推定 (estimation) のそれぞれが分離されている点が重要である。詳しくは、Hong (2015) あるいは VanderWeele (2015) などを参照されたい。

4 おわりに

因果媒介分析の理解には、頭の良さで性格の素直さが必要不可欠である！

参考文献

- [1] 阿部貴行・岩崎 学 (共訳) (2022). ローゼンバウム統計的因果推論入門. 共立出版.
- [2] 岩崎 学 (2015). 統計的因果推論. 朝倉書店.
- [3] Hong, G. (1997). Causality in a Social World: Moderation, Mediation, and Spill-over. Wiley.
- [4] VanderWeele, T. J. (2015). Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford University Press.

統計的因果推論の考え方と確率的介入に基づく因果効果

大阪大学大学院・理研 AIP 山本 倫生

1 はじめに

事象間の因果関係を解明することはあらゆる科学分野における目的の1つであろう。たとえば、医学研究において、ある疾患における新規薬剤の投与が疾患を改善させるかどうかを評価する場合、新規薬剤（原因）と疾患の改善（結果）の因果関係を検討することとなる。同様に、心理学研究において、ある尺度で測定される心理量の増加が抑うつ症状の程度を低下させるかどうかを評価する場合には、心理量（原因）と抑うつ症状の程度（結果）の因果関係を解明することが目的となる。一般に、データに基づいて統計学的に因果関係について推論を行うことを統計的因果推論とよぶ。ここで、医学研究の例において、新規薬剤を服用した個人に対する因果効果を考えてみよう。この個人に対する薬の因果効果は、薬を服用した場合（実際の状態）と飲まなかった場合（事実と異なる反事実の状態）とを比較することで定義される。しかし、両方の場合を同時に観測することはできないため、現実には個人に対する因果効果を適当な仮定なしに推定することは不可能である。そこで、統計的因果推論では個人に対する因果効果ではなく、集団に対する因果効果を定義し、データから推定する問題を考えることとなる。

2 統計的因果推論の考え方

ある処理変数（実験条件や治療を表す変数）について2つの水準があるとしよう。ルービン因果モデル（もしくはネイマン・ルービン因果モデル）では、ある処理変数 A の水準 a を割り付けた場合に観測されるであろう結果変数を表す記号 $Y^{A=a}$ （もしくは Y^a ）を導入する（たとえば Imbens and Rubin (2015) を参照）。これを潜在結果変数（potential outcome）という。前述のように、現実とは異なる処理の水準での潜在結果変数の値は観測できないため、個人に対する因果効果は推定できない。そこで、集団に対する因果効果を考えることとなる。たとえば、そのような因果効果としてよく利用されるものの1つに平均因果効果（または平均処置効果）がある。平均因果効果は「処理 $A = a$ を受けた場合に観測されるであろう結果変数 Y （つまり、 Y^a ）」と「処理 $A = a'$ を受けた場合に観測されるであろう結果変数 Y （つまり、 $Y^{a'}$ ）」との平均的な差 $\mathbb{E}[Y^a] - \mathbb{E}[Y^{a'}]$ で定義される。なお、記号 \mathbb{E} は期待値をとる作用素を表す。ここで、 $\mathbb{E}[Y^a]$ という量は、処理 $A = a$ を割り付けられた集団における期待値 $\mathbb{E}[Y|A = a]$ に必ずしも一致しないことに注意が必要である。つまり、いくつかの条件下でなければ、実際に観測されるデータから推定可能な $\mathbb{E}[Y|A = a]$ を処理の水準による群ごとに推定して比較しても、本当に知りたい潜在結果変数の期待値 $\mathbb{E}[Y^a]$ を比較していることにはならない。このことは、交絡が常に存在していると考えられる観察研究で特に注意が必要であり、適切な推定のためには、試験デザ

インの工夫や種々の調整解析が必要となる。なお、ランダム化比較試験の場合でも、平均因果効果を推定するためにはいくつかの条件の成立の吟味が必要である。

3 確率的介入に基づく因果効果

平均因果効果以外にも、実際の母集団と仮に介入を受けなかった場合の母集団の状態との比較を行う attributable effect や、共変量の値に依存して処理の水準の値を変化させる介入に基づく動的介入効果など、研究者の興味応じて様々な介入効果が利用可能である。一方で、これらの因果効果は母集団全体への確定的な割付に基づいて定義されており、このことが現実に即していない場合がある。たとえば、運動促進キャンペーンが健康に与える影響を評価する際に、母集団全体に毎日 30 分の運動を行ってもらおう介入を考えることはあまり意味がない（全く運動をしていない人にとっては運動促進となるが、もともと毎日 1 時間運動している人にとっては運動抑制となってしまふ）。この問題を解決する方法の 1 つに、各個人の現実での処理の水準の値に応じて割り付ける水準を変化させる介入を行うことが考えられる。このような介入は確率的介入 (stochastic intervention, または, modified treatment policy) とよばれ、たとえば Pearl (2009) では因果パラメータの導出が議論されている。このような確率的介入に関する研究が近年活発に行われており、より実情に即した因果効果の検討が進められている。発表当日は、確率的介入に基づく因果効果の推定について、Díaz and van der Laan (2012) および Haneuse and Rotnitzky (2013) に基づいて説明するとともに、近年の関連研究について紹介する。

参考文献

- [1] Imbens, G. W., Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.
- [2] Díaz, I., van der Laan, M. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics*, 68, 541-549.
- [3] Haneuse, S., Rotnitzky, A. (2013). Estimation of the effect of interventions that modify the received treatment. *Statistics in medicine*, 32, 5260-5277.
- [4] Pearl, J. (2009). *Causality*. 2nd Ed. Cambridge University Press.

エシェロン解析法：最近の展開と応用事例

岡山大学 環境生命自然科学学域 石岡 文生

1 はじめに

リモートセンシングデータや地域ごとの集計データなど、空間的位置を持つデータ構造の解析には、データの可視化が効果的である。最近では、GISをはじめとして様々な優れたツールが提供されているが、空間データ構造の客観的な表現は難しく、その解釈には主観的な側面も多く存在する。こうした問題に対して、エシェロン解析法は、空間的な位置を表面上のデータの高低に基づいて分割し、空間データの位相的な階層構造を系統的かつ客観的に視覚化する手法として開発された。特に、栗原(2003)によって最初に提案された「エシェロン解析法を用いたホットスポット（有意に高い値を示す領域群）の検出」に関する取り組みは、アルゴリズムの改良や実データへの応用、ソフトの開発など現在でも積極的な展開が続いている（Kurihara et al., 2020; 神原他, 2021; 梶西他, 2021; 竹村他, 2021; Takemura et al., 2022; 掃部他, 2023）。本報告では、エシェロン解析法を用いたホットスポット検出に関する話題を中心に、最近の展開について紹介する。

2 エシェロン解析法を利用したホットスポット検出

2.1 空間スキャン検定

「市区町村別の死亡数」などのように領域ごとに与えられたデータについて、「どこかにホットスポットは存在するのか？存在しているとしたら、その範囲はどこまでなのか？」という問いに、1つの解を与えてくれるのが空間スキャン検定である。これは、あるルールに基づいて領域を走査（スキャン）することにより、高い尤度を示す領域群（ホットスポット）を同定するもので、疫学分野などで広く利用されている。

いま、 m 個に分割された領域全体を \mathbf{G} とし、その \mathbf{G} 内で互いに隣接しあう任意の領域群をウィンドウと呼び \mathbf{Z} と表す。また、領域ごとに観測されるデータがある確率分布に従うと仮定し、その分布を定める1つの母数 θ に関心があるとす。ここで、あるウィンドウ \mathbf{Z} の中では母数 $\theta_{\mathbf{Z}}$ を持ち、そのウィンドウの外では異なる母数 $\theta_{\mathbf{Z}^c}$ を持つとき、帰無仮説を $H_0: \theta_{\mathbf{Z}} = \theta_{\mathbf{Z}^c} = \theta_{\mathbf{G}}$ (\mathbf{Z} はホットスポットではない)、対立仮説を $H_1: \theta_{\mathbf{Z}} > \theta_{\mathbf{Z}^c}$ (\mathbf{Z} はホットスポットである) とおく。このとき、空間スキャン統計量 λ は2つの仮説の下での尤度の比で定義され、次の \mathbf{Z} の関数として与えられる。

$$\lambda(\mathbf{Z}) = \frac{H_1 \text{ の下での尤度}}{H_0 \text{ の下での尤度}} = \frac{\prod_{i \in \mathbf{Z}} f(x_i | \theta_{\mathbf{Z}}) \prod_{i \notin \mathbf{Z}} f(x_i | \theta_{\mathbf{Z}^c})}{\prod_{i=1}^m f(x_i | \theta_{\mathbf{G}})}$$

ここに、 f は対象とするデータの分布を表し、例えば領域ごとに観測されるデータがカウントデータの場合はポアソン分布、成功数・失敗数で与えられるような2値データの場合は二項分布をあてることができる。なお、計算を簡単にするため、実際には λ を対数変換したものをを用いるのが一般的である。つづいて、ウィンドウ \mathbf{Z} の全体集合 $\mathcal{Z} = \{\mathbf{Z}_1, \mathbf{Z}_2, \dots\}$ の中から最大の対数尤度比をもつウィンドウ、すなわち、

$$\hat{\mathbf{Z}} = \arg \max_{\mathbf{Z} \in \mathcal{Z}} \log \lambda(\mathbf{Z})$$

なるウィンドウ $\hat{\mathbf{Z}}$ をホットスポットの候補と考える。 $\hat{\mathbf{Z}}$ が統計的に有意なホットスポットであるかどうかを評価するためには、 H_0 の下での $\max_{\mathbf{Z} \in \mathcal{Z}} \log \lambda(\mathbf{Z})$ の分布が必要になる。しかしながら、これを一意に定めることは解析的に困難なため、モンテカルロ法によるシミュレーションで求めた p 値によって有意性を評価するのが通例となっている。

2.2 エシェロンスキャン法

さてここで、現実的に計算可能な M 個の \mathbf{Z} の集合 $\mathcal{Z} = \{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_M\} \subset \mathbf{G}$ をどのように求めるか（これを「スキャンする」という）が問題になるが、栗原 (2003) によって提案されたのがエシェロン解析法によって \mathcal{Z} を求める方法である。エシェロン解析法 (Myers et al., 1997) は、近傍情報を与えられた領域上の 1 変量データ $h_i, i = 1, 2, \dots, m$ に対して、空間的な位置を表面上のデータ高低（濃淡）に基づき分割し、データの位相的な構造を系統的かつ客観的に見つける (図 1)。得られた構造を階層的に表現したグラフはエシェロンデンドログラムと呼ばれ、このグラフの上位の階層から順に、「階層を構成する領域」を上から順にスキャンしながら \mathcal{Z} を構築していく (図 2)。石岡・栗原 (2010) は、このスキャン方式をエシェロンスキャン法と命名するとともに、この方式によって任意の形状を持つ \mathbf{Z} の集合を構築でき、さらには他のスキャン方式と比べて計算コストが抑えられることを示した。

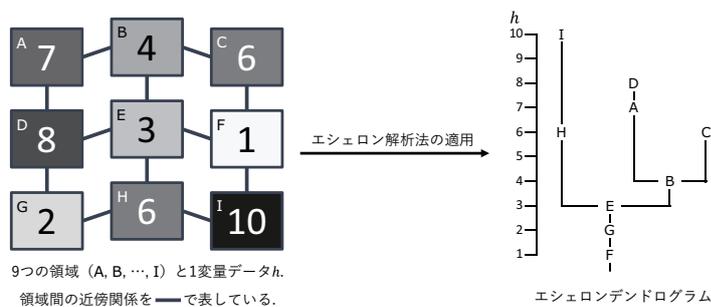


図 1: エシェロン解析法の適用例

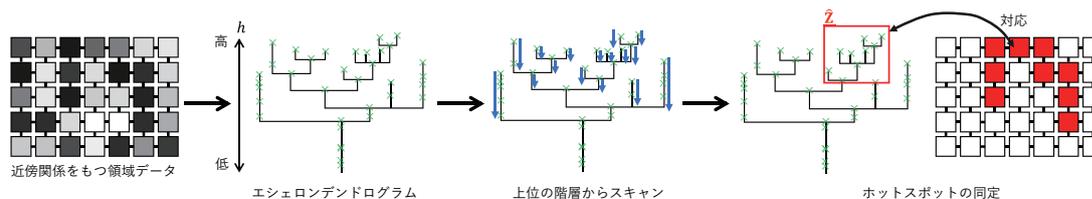


図 2: エシェロンスキャン法のイメージ

2.3 R パッケージ

エシェロンスキャン法の R パッケージ `echelon` (Ishioka, 2020) を紹介する。ここでは、栗原 (2003) で適用例として用いられた 1974 年 7 月～1984 年 6 月の米国ノースカロライナ州 100 郡における乳幼児突然死症候群のデータ (Sudden Infant Death Syndrome; SIDS) に対し、ホットスポットを検出するコード、ならびにその出力結果の一部とデンドログラム (図 3) を示す。有意なホットスポットとして 5 つの郡が同定されたことが見て取れる。

```

#spData パッケージからノースカロライナ州の SIDS データと近傍情報を取得
install.packages("spData")
data("nc.sids", package = "spData")
SIDS.cas <- nc.sids$SID74 + nc.sids$SID79 #SIDS 死亡数
SIDS.pop <- nc.sids$BIR74 + nc.sids$BIR79 #出生数
SIDS.exp <- SIDS.pop * sum(SIDS.cas) / sum(SIDS.pop) #期待死亡数
SIDS.rate <- SIDS.cas * 1000 / SIDS.pop #出生 1000 人当たりの SIDS 死亡数
ncCR85.nb #spData パッケージが提供する郡境界線に基づく近傍情報

#echelon パッケージを CRAN からインストールして呼び出す
install.packages("echelon")
library(echelon)

#h = SIDS.rate としてエシェロン解析法を適用
SIDS.echelon <- echelon(SIDS.rate, nb = ncCR85.nb, name = row.names(nc.sids))

#ポアソン分布に基づく統計量を用いたエシェロンスキャン
echepoi(SIDS.echelon, cas = SIDS.cas, ex = SIDS.exp, main = "Hgi rate clusters",
ens = FALSE, cluster.info = TRUE, K = 20) #ウィンドウ内の最大領域数は 20 に設定
text(SIDS.echelon$coord, labels = SIDS.echelon$regions.name, adj = -0.1, cex = 0.7)

```

```

MOST LIKELY CLUSTER -- 5 regions
Cluster regions included : Scotland, Hoke, Columbus, Bladen, Robeson
Number of cases .....: 139
Expected cases .....: 72.6694
Observed / expected .....: 1.9128
Relative risk .....: 2.0058
Log likelihood ratio ....: 25.3807
Monte Carlo rank .....: 1/100
P-value .....: 0.01

```

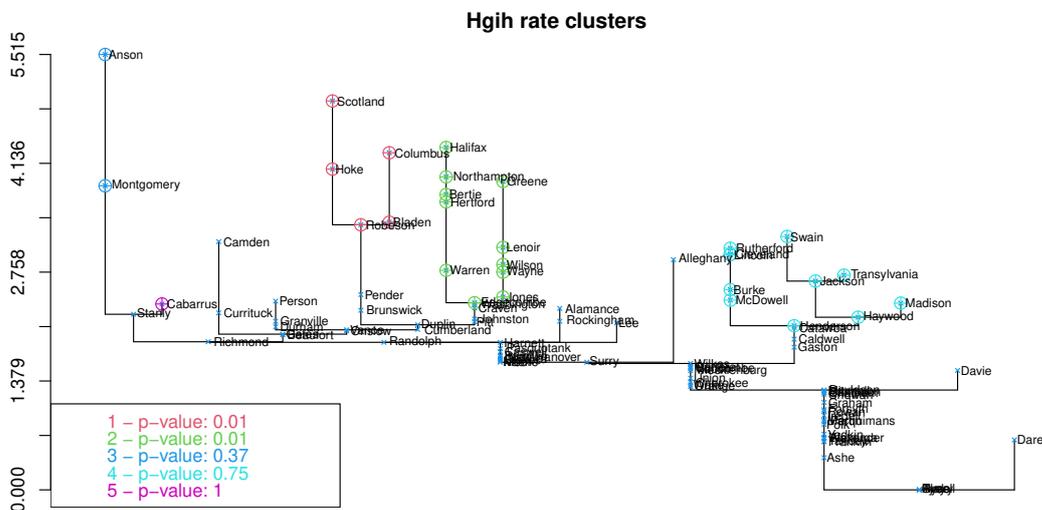


図 3: echelon パッケージを用いることによるエシェロンスキャンの結果

3 最近の展開

エシェロン解析法ならびにエシェロンスキャン法に関する最近の話題について紹介する。詳細については当日報告する。

3.1 可視化のためのアプリケーション

梶西他 (2021) は, R の shiny パッケージを用いて対話的な操作を行いながらエシェロンスキャンの結果を地図上に可視化する Web アプリケーションを開発しており, β 版を <https://fishi.ems.okayama-u.ac.jp/kajinishi/ver0.1> で公開している (図 4).

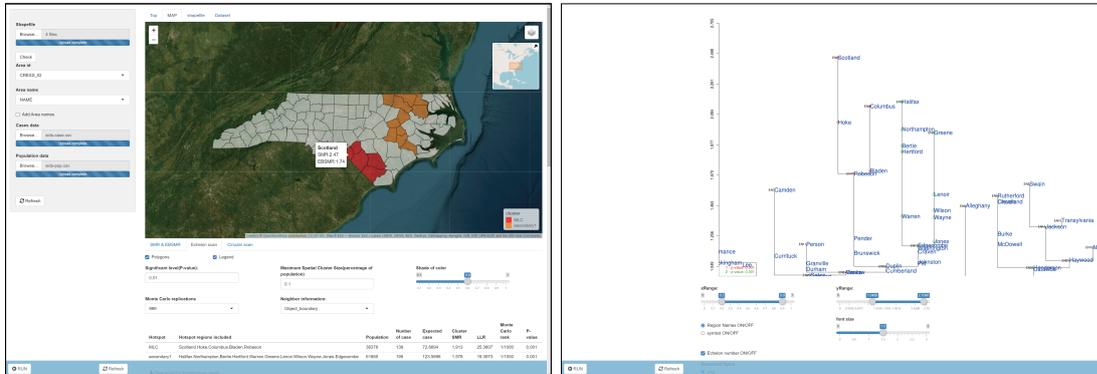


図 4: エシェロンスキャンのための Web アプリケーション

3.2 高リスクなホットスポット検出法の提案

エシェロンスキャン法は, 高尤度なウィンドウのスクランに注力するあまり, 低リスク領域 (例えば, ポアソン分布に基づく統計量を用いる場合に “実死亡数 < 期待死亡数” である領域) までをも含んだ, 巨大でいびつな形状をしたホットスポットを検出することがある. この問題に対処するため, 竹村他 (2021) や Takemura et al., (2022) では, エシェロンスキャン法に Tango (2008) の制限付きスクラン統計量を実装した Adjusted Echelon scan method (AESM) を提案した. これにより, 高リスク領域のみで構成されたホットスポットの検出を可能にし, さらには計算コストの大幅な削減に成功している.

参考文献

- [1] 石岡文生, 栗原考次. (2012). Echelon 解析に基づくスクラン法によるホットスポット検出について. 統計数理, **60**, 93–108.
- [2] Ishioka, F. (2020). *echelon v0.1.0: The Echelon Analysis and the Detection of Spatial Clusters using Echelon Scan Method*. <https://CRAN.R-project.org/package=echelon>
- [3] 梶西将司, 石岡文生, 栗原考次. (2021). 空間疫学におけるリスク評価とその可視化のためのソフトウェア. 計算機統計学, **33**, 91–105.
- [4] 掃部耀平, 竹村祐亮, 石岡文生. (2023). エシェロンスキャン法を用いた重み付き Normal モデルに基づくクラスター検出について. 日本計算機統計学会第 37 回シンポジウム講演論文集, 146–149.
- [5] 神原あん, 竹村祐亮, 石岡文生, 栗原考次. (2021). Prospective echelon scan による時空間クラスター検出手法の提案. 2021 年度日本分類学会シンポジウム, 2021 年 12 月.
- [6] 栗原考次. (2003). 階層的空間構造を利用したホットスポット検出. 計算機統計学, **15**, 171–183.
- [7] Kurihara, K., Ishioka, F. & Kajinishi, S. (2020). Spatial and temporal clustering based on the echelon scan technique and software analysis. *Jpn J Stat Data Sci*, **3**, 251–258.
- [8] Myers, W. L., Patil, G. P. & Joly, K. (1997). Echelon approach to areas of concern in synoptic regional monitoring. *Environmental and Ecological Statistics*, **4**, 131–152.
- [9] 竹村祐亮, 石岡文生, 栗原考次. (2021). Echelon scan 法による高リスクな空間クラスター検出法の提案. 計算機統計学, **34**, 23–43.
- [10] Takemura, Y., Ishioka, F. & Kurihara, K. (2022). Detection of space-time clusters using a topological hierarchy for geospatial data on COVID-19 in Japan. *Jpn J Stat Data Sci*, **5**, 279–301.
- [11] Tango, T. (2008). A spatial scan statistic with a restricted likelihood ratio. *Japanese Journal of Biometrics*, **29**, 75–95.

Covid-19 新規感染者数のデータ収集と可視化

東海大学理学部 山本 義郎

1 はじめに

新型コロナウイルス感染症が 2020 年から日本を始めとする世界各地で感染が広がった。東海大学のメインキャンパスがある神奈川県では市区町村ごとに感染者の公表がなされていたため、県内の感染状況について市区町村ごとの状況が把握できる可視化が可能であることがわかった。そのため統計ソフトウェア R のパッケージである Shiny を用いてインタラクティブなアプリケーションを実現し、棚橋 他 (2021) で発表を行った。その後、国土交通省の行政区域データを使用することで、コロプレスマップを作成できることが判明した。さらに 2021 年度には東海大学の学生及びキャンパス所在地周辺の地域住民の方を対象に周辺地域の感染状況を把握できるサイトを立ち上げた。

2022 年 9 月 26 日からは、ウィズコロナに向けた新たな段階へ移行するとして、市区町村ごとのデータ公開がなくなり神奈川県は保健所単位となり、更に 2023 年 5 月 8 日以降は定点医療機関からの 1 週間集計の報告となった。これらの変化への対応についても紹介する。

2 データについて

2.1 感染者数などのデータについて

本研究では、まずジャックジャパン株式会社から入手したデータ [2](更新が途中で止まったため、2020 年 11 月末まで) と、12 月以降は神奈川県や川崎市などの市区町村が公表した pdf 形式の資料からデータを取得した。横浜市については区ごとのデータは当初公開されておらず 1 週間の累計の感染者数が公表がされていたため、横浜市区別コロナデータを使用した。集計の際には記者発表日を用いて、居住市区町村ごとに感染者数の集計を行った。(2020 年 12 月以前は PCR 検査の陽性判明日)。また、可視化の際に人口を考慮するために、神奈川県の市区町村 (行政区も含む) の人口のデータ (2019 年 10 月の人口推計より) を使用した。また、集計したデータを地図に可視化するために、平成 31 年の行政区域データを使用した。路線図を表示する可視化の際には、令和元年の鉄道データを使用した。

2.2 データを取得するための処理

神奈川県の新型コロナウイルス感染症のオープンデータは市区町村別ではなく、保健所単位ごとだったが、政令指定都市や神奈川県の記者発表資料 (pdf) では、市区町村単位での記載があった。そのため統計ソフトウェア R を用いて、記者発表資料から神奈川県内の新型コロナウイルス感染症の感染者の情報を Web スクレイピングにより毎日自動的に取得

するようにした。まず、市区町村の pdf ファイルの URL を取得し、取得した URL をもとに PDF 上のテキストを取得し、文字列の切り出しなど文字列処理を行ない、整形・結合した。神奈川県以外のデータについても、同様に毎日自動取得するプログラムを作成した。

3 可視化の方法

3.1 コロプレスマップによるデータの可視化

可視化としては地図上で色の濃度によって数値を表現するコロプレスマップを用いた。まず統計ソフトウェア R のパッケージである Shiny を用いて、日付を指定した際に過去の状況もコロプレスマップとして確認できるインタラクティブな可視化を実現し、GitHub で公開した。次に Shinyapps.io を用いることで、R を利用しないユーザでも Web ブラウザから確認できるようにした。しかし、Shinyapps.io では、日本語が使用できず、アクセスが増えた場合の対応が不安だったため、一般には公開していなかった。その後、横浜市の区単位のデータを用い、横浜市の区ごとのコロプレスマップも作成した。横浜市の区別データは金～木の 1 週間ごとの感染者数になっており、他のデータとは、集計単位が異なるため、神奈川県の横に別途横浜市を表示することにした。次に、コロプレスマップに路線図を表示できるように変更した。しかし、R のライブラリである leaflet を用いると、表示までに時間がかかるため、R のライブラリの ggplot2 を用いて静止画として表示するように変更した。静止画の場合、路線図をつけると見にくくなるため、一般公開では路線図の表示をやめることにした。しかし、GitHub や Shinyapps.io では路線図の表示や leaflet を用いたアプリを公開していた。その後、簡単に日付を変えることができるように、前週、前日、翌日、翌週の 4 つのボタンを設置した。また、神奈川県全体の人数の合計の表示や、感染者の推移のタブなどの設置し、一般公開を行った。

3.2 折れ線グラフでの可視化

コロプレスマップとは別に、市区町村ごとの折れ線グラフが確認できるようにした。市区町村を複数選択し、選択したものについて折れ線グラフを表示するようにした。更に、重ね合わせの折れ線グラフを dygraphs パッケージ (dygraphs, 2018) を利用して作成し、カーソルを折れ線グラフに置くと対象の市区町村名がハイライトしたり、下部のバーをスライドさせることで表示期間を調整できるなどの機能を実装した。表示期間を調整した際に自動で最大値を調整し、感染者数の波を確認しやすくなる工夫を行った。

4 Shiny-Server による Web サイトの公開

このシステムを Web サイトとして一般公開するために、Linux 環境 (Ubuntu) の PC に Shiny-Server をインストールし、Shiny アプリケーションで作成した可視化システムをそのまま Web ページとして実装できるようにした。URL のドメイン名を取得し、2021 年 05 月 31 日に「リアルタイムな地域への感染症関連情報の提供サイト (<http://covid-map.bmi-tokai.jp/>)」を立ち上げた (図 1)。



リアルタイムな地域への感染症関連情報の提供サイト Real-time Local Information Provider

【概要】

Real-time Local Information Providerは、地域で役に立つ感染症関連情報をリアルタイムに提供することをめざす研究プロジェクト「地域へのリアルタイムな感染症関連情報の配信技術の開発」の成果を発信する場として運営されています。

【メニュー】

神奈川県市区町村ごとの新型コロナウイルス新規感染者数（インタラクティブなコロプレスマップ）

[PC版 / スマートフォン版](#)

東京都市区町村ごとの新型コロナウイルス新規感染者数（インタラクティブなコロプレスマップ）

[PC版 / スマートフォン版](#)

熊本県市区町村ごとの新型コロナウイルス新規感染者数（インタラクティブなコロプレスマップ）

[PC版 / スマートフォン版](#)

北海道の振興局ごとの新型コロナウイルス新規感染者数（インタラクティブなコロプレスマップ）

[PC版 / スマートフォン版](#)

静岡県の市区町村ごとの新型コロナウイルス新規感染者数（インタラクティブなコロプレスマップ）

[PC版 / スマートフォン版](#)

新規感染者数に基づく感染リスクの計算方法

いくつかの仮定を置いた上での議論ですが、[こちら](#)を参考にしてください。

[全国の都道府県ごとの新型コロナウイルス新規感染者数（PC推奨）](#)

[都道府県別 新型コロナウイルス感染者数の推移](#)

[ツイッター分析による新型コロナウイルス関連ニュース（準備中）](#)

[新型コロナ感染状況都道府県の状況報告（YouTube）](#)

リンク

他の都道府県について、類似した可視化を行っているサイトを紹介します。

・[沖縄県 COVID-19 概況](#)

最新の人口10万人あたりの新規感染者数（7日間累積）

※最新の公開区分（日付）は都道府県によって異なります。

2022-01-18 ~ 2022-01-24 (7日間累積)
神奈川県全体 255.53人



図 1: 当初の Web サイトでの公開

参考文献

- [1] Kulldorff, M. (2002). A spatial scan statistic. *Communications in Statistics: Theory and Methods*, **26**(6), 1481–1496.
- [2] Pebesma, E. (2018) Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439—446, <https://doi.org/10.32614/RJ-2018-009>
- [3] Gagolewski, M. (2022) stringi: Fast and portable character string processing in R. *Journal of Statistical Software*, 103(2), 1–59, <https://doi.org/10.18637/jss.v103.i02>
- [4] 棚橋真弓・山田実俊・今西規・山本義郎 (2021) 新型コロナウイルス感染症陽性者数可視化システムについて, 日本計算機統計学会第 35 回シンポジウム予稿集
- [5] dygraphs, Interface to 'Dygraphs' Interactive Time Series Charting Library, <https://cran.r-project.org/web/packages/dygraphs/>
- [6] ggplot2, Create Elegant Data Visualisations Using the Grammar of Graphics, <https://cran.r-project.org/web/packages/ggplot2/>
- [7] leaflet, Create Interactive Web Maps with the JavaScript 'Leaflet' Library, <https://cran.r-project.org/web/packages/leaflet/>
- [8] pdftools, Text Extraction, Rendering and Converting of PDF Documents, <https://cran.r-project.org/web/packages/pdftools/>
- [9] rvest, Easily Harvest (Scrape) Web Pages, <https://cran.r-project.org/web/packages/rvest/>
- [10] shiny, Web Application Framework for R, <https://cran.r-project.org/web/packages/shiny/>
- [11] ジャックジャパン株式会社. 都道府県別新型コロナウイルス感染者数マップ (2020) <https://gis.jag-japan.com/covid19jp/> (2020 年 12 月 1 日参照).
- [12] 国土交通省. 行政区域データ. 国土数値情報. https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N03-v2_4.html (2021 年 4 月 15 日参照)
- [13] 国土交通省. 鉄道データ. 国土数値情報. https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N02-v2_3.html (2021 年 5 月 7 日参照)