

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 21 日現在

機関番号：24402

研究種目：若手研究(B)

研究期間：2014～2016

課題番号：26730017

研究課題名(和文)カテゴリーカルデータ解析における新たなモデル及び統計量の開発

研究課題名(英文)Development of new statistical models in categorical data analysis

研究代表者

山本 紘司(Yamamoto, Kouji)

大阪市立大学・大学院医学研究科・特命准教授

研究者番号：10548176

交付決定額(研究期間全体)：(直接経費) 1,700,000円

研究成果の概要(和文)：本研究では、治療効果が「有効」「不変」「悪化」や「生存」「死亡」のようなカテゴリーカルな結果として表されるデータ解析において、既存手法よりも有用なアプローチについて研究を行った。上述のようなカテゴリーカルデータは分割表、もしくはクロス集計表と呼ばれる形式でまとめられることが多く、たとえば2つの薬剤間で効果が異なるのか、3つの地域で視力に差はあるのか、などを仮説としてデータ解析を行う。

具体的には、既存手法によるデータ解析よりもより詳細な、もしくはより正確な解析が行える新たな統計モデルの開発を行うなど、応用面でも貢献できる研究を行った。

研究成果の概要(英文)：Categorical data, which have some categories like "Response", "Stable" and "Progression" in evaluating a treatment, have been analyzed by using statistical models. Such data are summarized into "contingency tables".

In this research, I have proposed some new statistical models so that we can analyze such contingency table data in more detail.

These development is meaningful in not only mathematical aspect but application one.

研究分野：カテゴリーカルデータ解析

キーワード：分割表解析 診断研究 正確検定

1. 研究開始当初の背景

臨床試験では、アウトカムが順序カテゴリカル変数となることもままある。このときこれらのデータは分割表(クロス分類表)としてまとめることができる。分割表データの解析結果は、診療ガイドライン作成や医師の治療方針決定にも影響する可能性があることから、大きな意味を持つ。したがって、その解析手法が適切か否かが重要となる。順序カテゴリカルデータ(分割表)の解析においては、これまでに様々な統計モデルが開発されてきているが、モデルが複雑になりパラメータ推定に時間がかかるという課題もある。そこで、よりデータに適合するようなモデル開発および効率よくパラメータ推定を行えるような新たなモデル提案が次なる課題と考えられた。また、とくに順序カテゴリカルデータの解析では既存の適合度検定統計量は順序情報を利用していないため損をしている部分があり、この点をうまく利用した新たな統計量の開発も課題となる。さらに、名義カテゴリカル変数の経時データ解析においては適切な解析手法が未だ確立しておらず、この分野における新たな解析手法の構築は大きな意義をもつ。

2. 研究の目的

上述の学術的背景を踏まえ、本研究では次の3つを研究課題とする。

(1) 順序カテゴリカルデータ解析における累積モデルの直交分解定理を与える。

(2) カテゴリの順序を考慮した従来の統計量より検出力の高い統計量を開発する。

(3) 名義カテゴリカル変数に対する経時データ解析のためのモデル開発を行う。

また、各課題においては理論的研究のみならず、実データへ適用するなどの応用的研究も行う。

3. 研究の方法

本研究は、平成26~28年度の3年間で行う予定である。概ね1年ごとに、1) 順序カテゴリカルデータ解析における累積モデルの直交分解及び(従来の)適合度検定統計量の直交分解、2) カテゴリの順序を考慮した従来の検定統計量より検出力の高い適合度検定統計量の開発、3) 名義カテゴリカル変数に対する経時データ解析のためのモデル開発、の研究を行う。各課題について、具体的には下記のアプローチにより研究を遂行していく。

(1) 直行分解については、累積モデルに対するデザイン行列を定義し、その行列からなる線形空間と直交する空間(直交補空間)に存在するデザイン行列をもつモデルを探索することで累積モデルに対する直交分解を与える。

(2) 新たな検定統計量の開発については、帰無仮説及び対立仮説のもとで、改良した検定統計量の近似分布を導出し、従来の検定統計量の検出力との比較を行い、検出力の差の程度を評価する。

(3) 名義カテゴリカル変数の経時データ解析におけるモデル開発については、多項分布とロジットリンクを用い、あるカテゴリから別のカテゴリへ推移する推移確率を各カテゴリに対して計算し、それらを観測時点ごと時系列に計算する。さらに任意の1つのカテゴリを基準とした各カテゴリのオッズに対する介入効果を調べ、シミュレーションにより、新たな解析手法を適用した結果を検証する。

4. 研究成果

(1) 既存の累積モデルを拡張した新たな統計モデルを導入し、このモデルを用いた分解定理を与えた。また、当初予定にはなかったが、上記モデル開発を行う上で、新たな解析アプローチ(尺度)を見出したため、その尺度開発も行い、順序カテゴリカルデータ解析における新たな知見を得た。

(2) 及び(3)については、期間内に完全な成果を上げることはできなかった。ただし、現状で(3)に対する課題については推移確率を導入した統計モデルを構築し、提案手法の有用性をシミュレーションをとおして示した。こちらは現在論文執筆中である。

当初設定した課題については上述のとおりであるが、研究を進める上で見つかった課題も本研究における新たな課題として研究目的に追加した。具体的には、診断研究において、2つの検査法間でそれぞれ診断能を測り、その性能を比較することが目的の試験においては、アウトカムがカテゴリカルデータとなり分割表形式でデータをまとめることができる。このとき、診断能の評価指標としてよく用いられる「感度・特異度」については、2つの検査法間でそれらを比較する統計手法は確立しており、たとえばマクネマー検定などがよく用いられる。この診断指標は検査法そのものの性能を表すには適切であるが、診断を受ける側に立ったときの指標としては適切ではない。すなわち、検査で陽性と診断されたとき、その人が真に病気に罹患している確率がどの程度かを知りたい。そのような指標としては「陽性的中率・陰性的中率」という指標が用いられる。しかし、これらの中率の比較において、確立した手法は存在せず、手法論を述べた論文も5編程度しかない。さらに、これら既存の論文では、すべて大標本理論に基づいた議論がなされているため、サンプル数の少ない小規模臨床試験でこれらの手法を適用することが適切か否かが不明である。そこで、新たな研究課題として、ま

ず既存の手法が小規模臨床試験においてどの程度の性能をもっているのかを、シミュレーションをとおして検証し評価することを目的とした。次いで、サンプル数が少ない状況では従来の大標本理論を用いた手法は本来不適切であることが多いため、近似を用いない正確な手法を提案することをもう1つの目的とした。既存手法の性能評価及び、提案した正確な検定手法については、シミュレーション等を行い評価をし、こちらの結果を現在論文にまとめている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計6件)

Yamamoto, K., Aizawa, M. and Tomizawa, S. (2016). Measure of departure from sum-symmetry model for square contingency tables with ordered categories. *Journal of Statistics: Advances in Theory and Applications*, 16, 17-43. 査読有
DOI: 10.18642/jsata_7100121694

Yamamoto, K., Matsuda, Y. and Tomizawa, S. (2016). Collapsed double symmetry model and its decomposition for square contingency tables. *International Journal of Statistics and Probability*, 5, 31-37. 査読有
DOI: 10.5539/ijsp.v5n4p31

Yamamoto, K., Nakane, H. and Tomizawa, S. (2016). Symmetry model based on bivariate distribution for square contingency tables with ordered categories. *Journal of Statistics: Advances in Theory and Applications*, 15, 71-84. 査読有
DOI: 10.18642/jsata_7100121663

Yamamoto, K., Ohama, M. and Tomizawa, S. (2015). Vision data analysis based on new statistical models and decompositions. *Annals of Biometrics and Biostatistics*, 2, 1019-1025. 査読有
<https://www.jscimedcentral.com/Biometrics/biometrics-2-1019.pdf>

Yamamoto, K., Aizawa, M. and Tomizawa, S. (2015). Decompositions of sum-symmetry model for ordinal square contingency tables. *European Journal of Statistics and Probability*, 4, 12-19. 査読有
<http://www.eajournals.org/wp-content/uploads/Decompositions-of-Sum-Symmetry-Model-for-Ordinal-Square-Contingency-Table.s.pdf>

Yamamoto, K., Shimada, F. and

Tomizawa, S. (2015). Measure of departure from symmetry for the analysis of collapsed square contingency tables with ordered categories. *Journal of Applied Statistics*, 42, 866-875. 査読有
DOI: 10.1080/02664763.2014.993362

[学会発表](計8件)

相澤愛奈, 山本紘司, 富澤貞男. Measure of departure from sum-symmetry model for square contingency tables having ordered categories. 日本数学会, 2016/3/19, 筑波大学.

茂木嶺志, 田畑耕治, 山本紘司. 分割表における連関尺度に基づく周辺非同等性の指標. 日本統計学会. 2016/3/5. 東北大学.

掛札純平, 田畑耕治, 山本紘司. 順序カテゴリ正方形分割表における非対称モデルの提案. 日本統計学会. 2016/3/5. 東北大学.

松田裕也, 山本紘司, 富澤貞男. カテゴリを併合した正方形分割表における二重対称モデルと分解. 日本統計学会. 2016/3/5. 東北大学.

松田裕也, 山本紘司, 富澤貞男. 西方分割表における併合した二重対称モデルと分解. 日本計算機統計学会. 2015/11/27. まなぼつと幣舞.

Yamamoto, K. and Tomizawa, S. Symmetry models based on bivariate normal distribution for square ordinal tables and the performance. Eighth International Workshop on Simulation, 2015/9/24, University of Natural Resources and Life Sciences (Vienna, Austria).

安藤宗司, 山本紘司, 富澤貞男. 順序カテゴリ正方形分割表における累積確率に基づくモデルの分解. 応用統計学会. 2014/5/22, 統計数理研究所.

安藤宗司, 山本紘司, 富澤貞男. 正方形分割表における累積確率を用いた周辺非同等性に関する尺度. 日本計算機統計学会. 2014/5/18, 中央大学後楽園キャンパス.

6. 研究組織

(1) 研究代表者

山本 紘司 (YAMAMOTO, Kouji)
大阪市立大学大学院・医学研究科・特命准教授

研究者番号: 10548176

(2) 連携研究者

富澤 貞男 (TOMIZAWA, Sadao)

東京理科大学・理工学部・教授
研究者番号：50188778

新谷 歩 (SHINTANI, Ayumi)
大阪市立大学大学院・医学研究科・教授
研究者番号：00724395