

ランバス留学 研究成果報告

2025年 3月 18日

ランバス留学基金委員会 御中

所属：経済学部
職名：准教授
氏名：國濱 剛

以下のとおり、報告いたします。

研究制度	ランバス留学 (滞在国：アメリカ合衆国)
研究課題	発展途上国における人口学的情報のベイズ統計分析
研究実施場所	カリフォルニア大学サンタクルーズ校
研究期間	2023年 3月 6日 ～ 2025年 3月 5日 (24ヶ月)

◆ 研究成果概要 (2,500字程度)

上記研究課題に即して実施したことを具体的に記述してください。

本研究課題では、発展途上国における死因の分布推定を向上させるため、死因聞き取り調査データを用いた新しいベイズ統計モデルの開発に取り組んだ。特に、年齢と性別に応じた症状の関連性を考慮することで、従来の統計方法よりも高精度な死因の推定が可能になった。また、情報理論的指標に基づいて質問項目の重要度の数値化を可能とし、様々な症状と死因との相関関係についても分析を行った。これらの研究成果は、公衆衛生政策の策定や医療リソースの最適化に貢献するものであり、今後の応用が期待される。

研究の背景として、発展途上国では出生・死亡の登録制度が十分に整備されていない国が多く、死亡診断書の発行率が低いことが問題となっている。特に、農村部では医療機関のアクセスが限られており、自宅で死亡するケースが多数を占める。このため、死因を特定する公的記録がなく、疫学研究や医療政策の策定において大きな課題となっている。死亡の原因を正確に把握することは、感染症対策や慢性疾患の管理、公衆衛生施策の最適化に不可欠である。このような背景から、口頭剖検 (Verbal Autopsy, VA) という手法が近年広く用いられている。これは、死亡者の家族や介護者への聞き取り調査を通じて、死因を推定するものであり、公衆衛生研究において重要な役割を果たしている。しかし、VAデータはしばしば不完全であり、回答の曖昧さや偏りが含まれることが多い。従来の統計手法では、こうしたデータの欠損やノイズに対応することが難しく、分析の精度に限界があった。本研究課題では、VAデータの分析に適したベイズ統計手法を開発し、年齢や性別による症状の関連性を考慮した新しい因子モデルの提案を通して、症状と死因の関係をより正確に推定し、発展途上国の公衆衛生政策に貢献することを目指した。

本研究課題では、ベイズ統計モデルを用いた新しい分析手法を開発したが、具体的には、年齢や性別に依存する症状間の関連性を考慮することで、従来のモデルよりも死因の推定精度を向上させることを試みた。研究の進め方としては、まず、PHMRC (Population Health Metrics Research Consortium) の新生児死亡データを用い、医学的に診断された死因とVA調査データを統合し、調査項目をバイナリ変数として整理した。その後、症状、年齢、性別を考慮した確率モデルを定義し、既存の方法と比較しながらベイズ統計モデルを拡張した。特に、提案モデルでは症状の相関構造を年齢や性別の影響を受ける階層的構造で表現することで、より実際のデータ構造に即した推定を行うことを可能にした。また、VAデータでは質問項目が百個以上に及ぶことも多いため、質問項目の相関関係を表すパラメータを推定する際に、統計モデルがデータに過適合する問題が発生する恐れがある。これを避けるため、因子モデリングを応用した上でクロスバリデーション手法を用いて最適な因子数の選択を行い、VAデータが持つ相関構造を極力少ないパラメータ数で柔軟に表現できるようにした。さらに、情報理論的手法を用いて各症状の死因に対する影響度を定量化し、マルコフ連鎖モンテカルロ法を用いたベイズ推定による効率的なモデルのパラメータ推定方法を開発した。提案手法の有効性を実証するため、VA分野の一般的な統計手法であるInterVAやInSilicoVAに加えて、Naive BayesやRandom ForestなどのVA分野以外でも広く応用されている機械学習手法との比較分析を行った。

本研究課題の成果として、新しいベイズ因子モデルの開発により、従来の方法では考慮されていなかった年齢や性別による症状の違いを明確にモデルに組み込み、死因推定の精度を向上させた。具体的に、関連分野における一般的指標の原因別死亡率 (Cause-Specific Mortality Fraction, CSMF) に基づいて死因推定の精度を測り、PHMRCの新生児データを用いた実証分析において、様々なシナリオの元で提案手法が既存アプローチよりも優れたCSMFの正確度を示した。従来の統計手法では、死因の条件付きでの症状間の独立性や年齢・性別と死因の独立性など、実データの構造とは異なる仮定に基づいてモデルが構築されていたが、本研究のベイズ統計モデルでは、因子モデリングを通してデータが持つ相関構造を柔軟に表現することで高精度な死因推定が可能となった。また、相互情報量 (MI) などの情報理論的指標を応用することで、PHMRC新生児データの死因推定において重要と考えられる症状・人口学的情報の特定を行った。例として、出生児の体重、早産・過期産、年齢、性別などが死因と比較的強い相関関係を示すことが判明した。加えて、性別は直接的というより他の症状を通して間接的に死因に影響を与えるという特徴を明らかにした。これらの研究成果を学術論文としてまとめて応用統計学の国際雑誌に現在投稿中であり、肯定的なフィードバックを得ている。また、国際学会等での学術発表も併せて行ってきた。

今後の課題として、地理的要因の導入が挙げられる。本研究課題では年齢と性別に焦点を当てたが、地域ごとの環境要因が死因に与える影響も考慮すべきである。例えば、医療施設へのアクセスや地域ごとの栄養状況、生活環境の違いが死因に影響を与える可能性があるため、これらの要因を取り入れたモデルの構築が求められる。また、追加の人口学的要因の分析として、教育レベルや経済状況など健康に影響を及ぼす可能性のある要因をモデルに組み込む必要がある。さらに、調査の簡略化と効率化に向けて、MIを活用しVA調査の質問項目を最適化することで、負担を軽減しながら精度を維持する方法を模索する。また、より広範なデータセットを用いたモデルの検証を行うことも重要である。本研究課題ではVA分野において一般的なPHMRCの新生児データセットを用いた実証分析を行ったが、他にも幼児、成人データセットが存在することに加え、近年PHMRCとは異なるVAデータも様々な発展途上国において新たに収集されている。異なるVAデータを用いた実証分析における提案手法のパフォーマンスを精査することで、見落とししていた欠点や今後の拡張の方向性が明らかになることが期待される。

以上