氏名（本籍）　安藤宗司（北海道）
学位の種類　博士（理学）
学位記番号　甲第1111号
学位授与の日付　平成28年3月18日
学位授与の要件　学位規則第4条第1項該当
学位論文題目　**Model and Measure of Symmetry for Ordinal Square Contingency Tables**
（順序正方分割表における対称性のモデルと尺度）

論文審査委員　（主査）教授　富澤　貞男
教授　明石　重男　　准教授　宮本　暢子
教授　立川　篤　　教授　瀬尾　隆

論文内容の要旨

カテゴリカルデータの統計解析手法について，20世紀初頭より多くの論文において議論がされてきた．また，カテゴリカルデータの統計解析手法の発展により，医薬，疫学，教育学，社会科学など幅広い分野で用いられている．

カテゴリカル変数が2つのカテゴリを持つとき，二値変数と呼ぶ．たとえば，“はい—いいえ”または“成功—失敗”などが該当する．カテゴリカル変数が2つ以上のカテゴリを持つとき，変数を2つのタイプに区別する．カテゴリ間に自然な順序を持たない変数を名義変数と呼ぶ．たとえば，音楽の種類（“クラシック”，“カントリー”，“フォーク”，“ジャズ”，“ロック”）などが該当する．カテゴリ間に自然な順序を持つ変数を順序変数と呼ぶ．たとえば，裸眼視力（“良い”，“やや良い”，“やや悪い”，“悪い”）や値の水準（“上位”，“中位”，“下位”）などが該当する．このカテゴリカル変数の区別は，データ解析する際にどの統計的手法が適切であるかを選択する際に重要となる．

数理統計学における分割表について考える．たとえば，ある人間の集団についてタバコを吸うか吸わないかと肺癌であるかないかによって4つの集団に分類することができる．各集団の人数からなる2×2型の表を数理統計学では2×2分割表を呼ぶ．同様に，それぞれr個のカテゴリとc個のカテゴリから構成されているときは，r×c分割表と呼ぶ．たとえば，左右の裸眼視力を良い・やや良い・やや悪い・悪いと4つのカテゴリに分けるとすると，4×4分割表を得ることができる．また，このようにr個のカテゴリとc個のカテゴ

リの 2 種類が同じ分類からなる分割表を正方分割表と呼ぶ.

実際の分割表解析において，我々は観測度数しか得ることができない．しかし，我々の関心は，観測度数ではなく，その観測度数の背後にある未知の確率分布を高い信頼度で推測することにある．そして，変数間の相互関連性に関して，有用かつわかりやすい解釈を実際のデータに対して与えることが分割表解析の目的である．そのために，データに良く適合し，解釈が容易である未知の確率分布に対する統計モデルを導入する必要がある．さらに，モデルにおける未知パラメータの推定法やモデルの適合の良さを計る検定統計量の開発，モデルの分解，複数の分割表データに対してモデル適合の良さを比較することができる尺度の開発など解決すべき多くの課題が存在する．本論文は分割表の中でも，正方分割表において，これらの課題を取り扱った研究である.

分割表解析において，多くの人の関心は，行変数（r 個のカテゴリ）と列変数（c 個のカテゴリ）が独立かどうか（関連性がないかどうか）にある．しかし，2 種類が同じ分類からなる正方分割表の特徴として多くの場合，観測度数が分割表の主対角セル付近に集中するため，行変数と列変数の相互関連性は極めて強く，統計的独立性は成り立たない傾向がある．そこで，正方分割表においては統計的独立性に代わって，対称性や非対称性に関する統計モデルが提案され，解析が行われてきた．1948 年に分割表における未知のセル確率の対称性の構造を示す対称モデルが導入された．その後，対称モデルを拡張したモデルとして様々なモデルが導入されている．対称性の構造を示すモデルとして，周辺同等モデルや準対称モデルが導入された．非対称性を示すモデルとしては，条件付き対称モデル，対角パラメータ対称モデル，線形対角パラメータ対称モデルなどが提案された．また，「対称モデルが成り立つための必要十分条件は，準対称モデルと周辺同等モデルの両方が成り立つことである」という対称モデルの分解定理が与えられている．対称モデルの分解定理は，与えられたデータに対して対称モデルの適合が良くない場合，適合が良くない原因を詳細に解析するのに有用である.

また分割表解析において，モデルの適合が良くない場合，対応するモデルからの隔たりの程度を測ることにも関心がある．1994 年に対称モデルからの隔たりを測る尺度が提案された．尺度は，複数の分割表データにおいてモデルからの隔たりの程度を比較するのに有用である.

本論文は 4 つの章から構成されており，順序正方分割表において，非対称性に関するモデルの提案や分解定理，周辺同等モデルを拡張したモデルからの隔たりを測る尺度の導出を扱っている.

第 1 章ではカテゴリカルデータの研究に関する歴史的背景の中でも特に正方分割表の研究と各章の概要について述べた.

第 2 章では，分割表の主対角線に対して非対称な構造を示す，リジットスコア型 2 パラメータ対称モデルを導入した．分割表データに対して、カテゴリが等間隔に区切られていない場合には，各カテゴリにスコアを導入したモデルを用いて解析することが適切であると考えられる．導入したモデルは，既存の非対称性に関するモデルに未知のリジットスコアを導入したモデルとなっている．また，導入したモデル，セル確率の和の対称性の構造

を表すモデル，リジットスコアを用いた平均一致の構造を表すモデルを用いて，対称モデルの分解定理を与えた．この分解定理は，実際のデータ解析において対称モデルの当てはまりが悪いとき，その原因を導入したモデルを用いてより細かく考察することに有用であると考えられる．そして，導入したモデルと分解定理から有用な結果と解釈が得られることを実データ解析とともに示した．

第3章では，平均周辺同等モデルからの隔たりを測る尺度を導入した．導入した尺度は，複数の分割表データにおいて平均周辺同等モデルからの隔たりの程度を比較するのに有用である．順序カテゴリ正方分割表において，周辺同等モデルは，行変数の周辺分布関数と列変数の周辺分布関数が同等であるという構造を表す．実際のデータ解析において，周辺同等モデルが成り立たないとき，我々は行変数の周辺分布関数と列変数の周辺分布関数にどのような違いがあるのかに興味がある．周辺同等モデルからの隔たりは，「行変数の周辺分布関数が列変数の周辺分布関数より大きい」または「行変数の周辺分布関数が列変数の周辺分布関数より小さい」という異なる 2 種類の方向性がある．順序カテゴリ正方分割表において周辺同等モデルからの隔たりを測る尺度が 2003 年と 2006 年に提案された．しかしながら，これらの尺度は周辺同等モデルの隔たりについて異なる 2 種類の方向性を区別することができない．したがって，第 3 章では異なる 2 種類の方向性を区別できる尺度を導入した．導入した尺度は，周辺同等モデルからの隔たりを直接的に測ることはできない．しかし，分割表データに対して，拡張周辺同等モデルの適合が良い場合，導入した尺度は周辺同等モデルからの隔たりを測ることが可能となる．導入した尺度の有用性について，理論面だけではなく実データを用いた解析例とともに示した．

第4章では，一般化周辺同等モデルからの隔たりを測る尺度を導入した．導入した尺度は，複数の分割表データにおいて一般化周辺同等モデルからの隔たりの程度を比較するのに有用である．一般化周辺同等モデルは，拡張周辺同等モデルをさらに拡張したモデルである．一般化周辺同等モデルをさらに拡張したモデルと導入した尺度の関係についてシミュレーションを実施し，導入した尺度が一般化周辺同等モデルからの隔たりを適切に測ることが出来ているかを考察した．その結果，導入した尺度は一般化周辺同等モデルからの隔たりを適切に測ることが示された．既存の尺度では一般化周辺同等モデルからの隔たりを測ることはできなかったため，導入した尺度により，一般化周辺同等モデルからの隔たりを測ることが可能となった．さらに導入した尺度の有用性について理論面だけではなく実データを用いた解析例とともに示した．

論文審査の結果の要旨

統計学における分割表解析は，教育学，心理学，社会学，医学，薬学，理学，工学など幅広い分野で利用されている．医薬分野における新薬と偽薬の比較試験において，新薬による改善度が，改善，やや改善，不変，そして偽薬による改善度が，改善，やや改善，不変によって 9 つの小集団に分けると，各集団の患者数からなる 3×3 分割表を得ることがで

きる．このような２種類の同じ分類からなる分割表を正方分割表と呼んでいる．特にカテゴリに順序がある場合は順序正方分割表という．

分割表解析において，我々は全体（母集団）の一部分として得られた標本に基づく観測度数しか得ることができない．我々の関心は，母集団における未知の確率分布がどのような構造になっているのかを（母集団の一部分として）得られた観測度数から，できる限りわかりやすい解釈が得られるように，高い信頼度で推測することにある．そのためには，データに良く適合し，かつ解釈が容易な確率分布に関する統計モデル（仮説）を導入する必要がある．さらに，モデル間の関連性，モデルの適合度を調べるための検定統計量の開発，モデルからの隔たりを測る尺度の開発など，多くの解決すべき問題点がある．

一般の２元分割表の解析において，多くの人の関心の一つは，分類間が独立（関連性がない）かどうかである．しかし，同じ分類からなる正方分割表の解析においては，分類間の関連性は極めて強く，統計的独立性は成り立たない．代わって，分類間の対称性に関心がある．たとえば，人間は，幼児の頃は左右裸眼視力は対称的であるが，成長とともに対称性は崩れる傾向にあり，左右どちらが良い目なのか，また，どのように非対称になるのかに関心がある．そのために，対称性あるいは非対称性に関する統計モデルを用いた解析が行われる．歴史的には，1948 年に Bowker が対称モデルを導入したのが始まりである．その後，対称モデルを拡張したいくつかのモデルが提案されている．

本論文は，カテゴリに順序のある同じ分類からなる正方分割表において，非対称モデルの提案，モデルの分解そしてモデルからの隔たりの程度を測る尺度を提案している．本論文は，４つの章から構成されている．

第１章は，対称性や非対称性のモデルと尺度の紹介や歴史的背景を述べている．

第２章では，正方分割表において，リジットスコアを用いた非対称モデルを提案している．このモデルは対称モデルの拡張である．また，「対称モデルが成り立つための必要十分条件は，提案したモデル，グローバル対称モデルそして周辺リジット一致モデルの全てが成り立つことである」という定理を与えている．この分解定理は，特に，データ解析において，対称モデルが成り立たない場合，その原因を詳細に解析するのに有効である．そして，英国の女性の左右裸眼視力データを用いて，新しいモデルと導出した分解定理の有用性を示している．

第３章では，平均周辺同等モデルを提案し，そのモデルからの隔たりを測る尺度を提案している．提案された尺度には三角関数を用いた独創的なアイデアがうまく導入されている．周辺同等性が成り立つとき尺度は 0 となり，行変数が列変数より大きくなる確率が 0 のとき尺度はプラス 1 になり，逆の場合は尺度はマイナス 1 となる．よく知られた相関係数は２つの変数間の独立性からの隔たりを測る尺度であるが，本論文で提案された尺度は周辺同等性からの隔たりを測る尺度である．提案された尺度の値は未知であるが，尺度の推定法や推定尺度の近似分散，そして尺度の信頼区間を理論的に与えている．さらに実際のデータを用いて提案した尺度の有用性を示している．

第４章では，一般化周辺同等モデルからの隔たりを測る尺度を提案している．これは Kullback-Leibler 情報量を含む power-divergence を用いた独創的な尺度である．そして，

ある一般化した周辺同等モデルの下でのシミュレーション実験により提案した尺度の妥当性を示している．さらに，具体的応用例を示し，提案した尺度の有用性を示している．

以上，本論文は理論面と応用面の両面において大変高く評価できるものであり，分割表統計解析の分野に，独創的な新しい解析法を与えており，この分野に大きな貢献をしている．よって理学的に価値ある知見と成果を得たもので博士（理学）の学位論文として十分価値あるものと認める．