Wonderful R［5］巻　統計的因果推論の理論と実装―潜在的結果変数と欠測データ―
高橋 将宜著･石田 基広監修･市川 太祐･高橋 康介･高柳 慎一･福島 真太朗･松浦 健太郎編
https://www.kyoritsu-pub.co.jp/bookdetail/9784320112452

はじめに

本書のタイトルは，『統計的因果推論の理論と実装：潜在的結果変数と欠測データ』である．因果とは，原因と結果の関係である．そのような因果関係は目に見えないため，データから推し測って考えることが本書の主題である．理論とは統計的因果推論の数理的メカニズムを意味しており，実装とは統計環境 R による数値解析を意味している．本書は，統計的因果推論の理論（数理的メカニズム）と実装（R による数値解析）の両方を統一的にカバーしたものである．副題の潜在的結果変数とは，ハーバード大学統計学科の Donald B. Rubin の提唱した潜在的結果変数の枠組みによる統計的因果推論を意味している．また，副題の欠測データとは，データの一部が観測されない場合の因果推論も扱うことを意味している．

本書の対象

理論とは統計的因果推論の数理的メカニズムを意味していると述べた．しかし，本書の数理的な理論解説は，できるだけ高校数学の範囲内で理解できるように工夫している．微積分や線形代数も，ほぼ登場しない．さらに，必要な数学的知識は，登場する箇所で解説もしている．また，本書では，統計的因果推論の数学的な理論を理解するだけでなく，R を用いた数値解析によってその理解を深め，実際にデータから因果推論を行う考え方と技術を学ぶ．

そこで，本書における R の使用には，2 つの目的がある．1 つ目は，数学が苦手な人にも，R を使った数値計算により，統計的因果推論のメカニズムを理解してもらうことである．この際，R の初心者にも数式と R コードとの対応関係がわかってもらえるように，できるだけ 1 行ごとに完結するコードを書くように心がけている．2 つ目は，R を使った統計的因果推論の実証研究ができるようになることである．本書の解析結果は，シミュレーション結果を除いて，すべて，本書の中に記載されている R コードを使って再現できるようにしている．

本書は，統計学を一度でも学んだことのある人を想定している．大学の授業であったり，入門的な書籍であったり，その方法はいくつかあるだろう．統計学を十分にマスターしている必要はないが，統計学の授業や書籍で「相関は必ずしも因果を意味しない」と指摘されたことを覚えている人が対象という意味である．つまり，相関係数が 1.0 または −1.0 に近かったとしても，必ずしもそれは因果関係があることを意味しないといわれたはずである．しかし，因果関係を明らかにすることは，多くの実証研究における目的である．「相関は必ずしも因果を意味しない」のであれば，因果推論をするためには，どうすればよいのだろうか？　本書は，相関係数のような統計学の基本的な用語を知っているが，データからどのようにして因果推論を行えばよいかについて知

りたい読者を対象としている．

たとえば，国ごとのチョコレート消費量とノーベル賞受賞者数には正の相関があることが知られている．ここから，チョコレート消費量の多い国は，ノーベル賞受賞者数を多く輩出する傾向があるといえる．しかし，チョコレート消費量を増やすと，ノーベル賞受賞者数を増やすことができるという因果的な議論は可能だろうか？　これについては，本書の第 6 章で実際にデータを解析して解決する．

本書のサポートページ

共立出版の本書サポートページは，`https://www.kyoritsu-pub.co.jp/bookdetail/9784320112452` である．本書は，誤植などがないように入念なチェックを繰り返しているが，それでも出版後に誤植が見つかる可能性がある．そのような誤植に関する情報は正誤表にまとめ，こちらに掲載する．また，本書の解析に使ったすべてのデータは，筆者の GitHub のサポートページ `https://github.com/mtakahashi123/causality` からもダウンロードできる．GitHub のサポートページからは，Code をクリックして，Download ZIP をクリックすればよい．

本書の構成

本書は，全体で 21 章から構成されている．本書は，第 1 章から順番に読んでいくことで，知識と技術が積みあがっていくように構成している．以下に概略を述べておこう．

第 1 章～第 9 章は，基礎的な考え方を扱っている．ここでは，統計的因果推論の考え方が凝縮されている．仮定が満たされているとき，重回帰モデルがどのようなメカニズムで交絡を取り除いているかを理解することは，統計的因果推論の立場から，本質的に重要である．したがって，重回帰モデルは因果推論に使えないなどとは思わずに，そのメカニズムをしっかりと押さえてほしい．

第 10 章～第 18 章では，傾向スコア，操作変数法，回帰不連続デザインといった統計的因果推論の実践的な技術を扱っている．それぞれの手法について，理論的なメカニズムの話から，最先端の議論までカバーし，具体的に R で解析する方法も解説している．実証研究において，解析対象になっているデータおよび知りたい推測対象の特性に応じて，これらの手法を使い分けてほしい．

第 19 章～第 21 章では，欠測データ解析と統計的因果推論の関連性について扱う．潜在的結果変数の枠組みでは，因果推論は欠測データの問題といわれるが，欠測データ解析は統計的因果推論に関する多くの書籍では，実際には扱われてこなかった事項である．

本書の内容

それぞれの章の内容について，もう少し詳しく説明しておこう．

第 1 章～第 4 章では，統計的因果推論の基本的な考え方を導入している．第 1 章では，因果推論の考え方に慣れるために，いくつかの日常的な具体例を紹介している．第 2 章では，具体的な数値例とフォーマルな理論を通じて，潜在的結果変数の枠組みを導入する．第 3 章では，潜在的結果変数の枠組みによって統計的に因果を推論するために必要な仮定を導入する．第 4 章では，推測統計の基礎的な事項の復習を扱っている．標準誤差と信頼区間は，統計的因果推論に特有なものではないが，推測統計における非常に重要な考え方であり，つまずきやすいポイントであるから，1 つ

の章を割いて解説している．標準誤差や信頼区間は，本書でも実際によく登場するので，ここで押さえておいてほしい．

第 5 章～第 9 章では，回帰モデルを扱う．回帰モデルは，統計的因果推論の最も基本的な考え方と技術であるから，これを理解することは必須事項である．第 5 章では，一次関数から始めて，通常の最小二乗法のメカニズムと単回帰モデルの基礎を与える．第 6 章では，バレンティン・ベン図を使って，重回帰モデルで交絡を統制することの意味を具体的かつ視覚的に解説する．第 7 章と第 8 章では，ややテクニカルな話題が続くが，重回帰モデルにおいて必要となる仮定を確認する．特に，7.3 節は，観察研究から統計的因果推論を可能にするための議論であるから，統計的因果推論の立場から極めて重要である．第 9 章では，前半において交互作用項のある重回帰モデルを扱う．また，9.3 節では，観察研究から統計的因果推論を行う際に，どのような共変量をモデルに含めるべきか議論している．9.3 節の議論も，統計的因果推論の立場から極めて重要である．

第 10 章～第 12 章は，傾向スコアを扱う．第 10 章では，傾向スコアとは何かを扱い，その性質について解説する．第 11 章では，観察研究における統計的因果推論手法としてよく使用される傾向スコアマッチングを扱う．マッチングのメカニズムから具体的なアルゴリズムの解説まで扱っている．また，近年，「傾向スコアをマッチングに使うべきでない」という主張がされたが，これについても検討している．第 12 章では，傾向スコアによる層化解析と重み付け法を扱っている．それぞれ，標本調査における手法と関連付けて，直感的に理解できるように工夫している．ここまで学んだあと，もう一度，9.4 節と 9.5 節を読んでみるとよいだろう．傾向スコアと重回帰モデルの相違点が，はっきりとするはずである．

第 13 章と第 14 章では，操作変数法を扱う．第 13 章では，再びバレンティン・ベン図を使って，なぜ未観測の交絡があっても，操作変数法によって適切な因果推論を行うことができるのか，具体的に考察する．第 14 章では，操作変数法の応用手法として，無作為化奨励デザインを扱う．これは処置の割付けが守られていない場合の因果推論手法であり，第 18 章とも関連がある．

第 15 章～第 18 章では，回帰不連続デザインを扱う．回帰不連続デザインは，近年，特に注目されている手法であるが，和書・洋書を問わず，具体的な解説は少ない．第 15 章では，回帰不連続デザインのメカニズムについて，図や数値例を使って，具体的かつ直感的に解説する．第 16 章では，実データを用いて，回帰不連続デザインのメカニズムを深堀りして検討する．また，第 16 章の後半では，カーネル関数，バンド幅の選択，RD プロットなど，回帰不連続デザインに関わる技術的な事項の解説をする．第 17 章の前半では，回帰不連続デザインによる統計的因果推論を可能にするために必要な仮定を扱う．また，第 17 章の後半では，ここまでに学んだことを総動員して，回帰不連続デザインによる実データ解析を行う．第 18 章では，ファジーな回帰不連続デザインを扱う．これは，局所的な範囲で操作変数法を実行するものであり，回帰不連続デザインではあるものの，第 14 章との関連も強い．

第 19 章～第 21 章では，欠測データ解析と統計的因果推論の関連性について扱う．第 19 章では，データが欠測しているときにどうすればよいかについて論じる．筆者はすでに，2017 年に『欠測データ処理』という書籍を執筆しており，この話題については 1 冊の本を書いた．本書の第 19 章は，『欠測データ処理』と対応させながら読むとよい．また，19.10 節は，『欠測データ処理』では扱わなかった交互作用項のあるモデルにおける欠測値の処理を扱っている．第 20 章では，傾向スコアマッチング，操作変数法，回帰不連続デザインにおいて，データが欠測しているときに，どのようにしてモデリングすればよいかを論じる．これは，和書・洋書を問わず，本書だけの特色ある

内容である．最後に，第 21 章は，最先端の話題である．本書で扱う Rubin の潜在的結果変数の枠組みでは，因果推論は欠測データの問題といわれる．しかしながら，通常，統計的因果推論と欠測データ解析は，別々の学問として発展してきた．第 21 章では，欠測値の処理手法である多重代入法を統計的因果推論の手法として活用する方法について論じる．

カバーイラストについて

本書のカバーイラストについて，簡単に説明しておこう．本書のテーマは，Rubin により提唱された潜在的結果変数の枠組みによる統計的因果推論である．表紙のイラストは，以下の反実仮想（潜在的結果変数）をイラストで表現したものである．

事実：太陽があるとき，喜ぶ．
反事実：曇っているとき，残念がる．ただし，反事実は観測されない（斜線）．

謝辞

石田基広先生（徳島大学）には，本書の原稿をお読みいただき，R コードや文章についてご監修いただいた．特に，R コードについて細かくチェックしていただき，可読性を向上させることができた．編集委員の先生方にも本書の原稿を閲読していただいた．大谷早紀氏（共立出版編集部）には，本書の企画段階から校正にいたるまで，大いにお世話になった．

佐藤俊太朗先生（長崎大学病院）には，生物統計学の専門家の視点から，本書の原稿を隅々までご確認いただいた．佐藤先生からは，数えきれないほどの有益なコメントをいただき，本書を大幅に改善することができた．矢内勇生先生（高知工科大学）には，第 18 章の執筆の際にご相談にのっていただいた．

本書の内容の一部は，長崎大学情報データ科学部の第 4 回コロキウム（2021 年 3 月 17 日）にて報告し，コメントをいただいた．また，本書の内容の一部は，立教大学，東洋大学，東京外国語大学，鹿児島国際大学，鹿児島大学，長崎大学にて筆者の担当した統計学関連の講義資料にも基づいている．受講者の中には鋭い質問をしてくれた学生もおり，それらは間接的に本書の改善へとつながっている．

最後になったが，本書の内容は，筆者の博士論文の指導教員である岩崎学先生（統計数理研究所）から学んだことがベースになっている．岩崎先生にも本書の原稿をご確認いただいたところ，「よく書けていると思いました．この本が早く出版されることを望んでいます．」と激励をいただいた．岩崎先生のご著書『統計的因果推論』（朝倉書店）も一緒に読んでいただくと，本書との相乗効果で，統計的因果推論の理解が深まるであろう．

このように，本書は，多くの方々からのご支援の末に出版することができた．この場にて，各位に深く謝意を表したい．

2021 年 11 月

高橋 将宜