

目 次

日本語版への序文	i
Stata の便利なコマンド集	xi
R の便利なコマンド集	xiv
序 文	xvii
学生のみなさんへ——統計学を学ぶ皆さんはこの本から何を学べるのか	
教師の皆さんへ——統計学を学生にどのように教えればいいのか	
謝 辞	xxxiv

第 1 章 因果関係の探求 ————— 1

1.1 特定化——基本となるモデル	3
1.2 2つの大きな課題——ランダム性と内生性	10
ケーススタディ インフルエンザの予防接種	18
ケーススタディ カントリー・ミュージックと自殺	22
1.3 絶対的基準としてのランダム化実験	26
結 論	33
キーワード	34

第 2 章 現場で使う統計 ————— 35

——データの適切な扱い方

2.1 データを知る	37
2.2 再現性	40
ケーススタディ アメリカの暴力犯罪	44
2.3 統計ソフトウェア	46
結 論	47
文献案内	48
キーワード	48
プログラミング・	
コーナー	49
演習問題	58

第 I 部 OLS の分析枠組みを理解する

第 3 章 二変量 OLS ————— 65
——計量社会科学分析の基礎

3.1	二変量回帰モデル	67			
3.2	係数推定値におけるランダムなばらつき	74			
3.3	内生性とバイアス	81			
3.4	推定値の精度	86			
3.5	確率の限界と一貫性	90			
3.6	解答可能な問題——不均一分散と誤差同士の相関	94			
3.7	適 合 度	98			
	ケーススタディ 身長と賃金	102			
3.8	外 れ 値	106			
結 論	110	文 献 案 内	111	キ ー ワ ー ド	112
コーナ	112	演 習 問 題	119	プ ロ グ ラ ミ ン グ ・	

第 4 章 仮説検定と区間推定 ————— 125
——リサーチクエスションに答える

4.1	仮 説 検 定	127			
4.2	t 検 定	134			
4.3	p 値	144			
4.4	検 出 力	146			
4.5	仮説検定に関する率直な説明	154			
4.6	信 頼 区 間	156			
結 論	160	文 献 案 内	161	キ ー ワ ー ド	161
コーナ	162	演 習 問 題	164	プ ロ グ ラ ミ ン グ ・	

第 5 章 多変量 OLS ————— 169
——中心となる手法

5.1	多変量 OLS を用い、内生性に対処する	171			
5.2	欠落変数バイアス	181			
	ケーススタディ 教育は経済成長につながるか	185			

5.3	測定誤差	189			
5.4	精度と適合度	192			
	ケーススタディ 制度と人権	200			
5.5	標準化係数	203			
5.6	複数の係数に関する仮説検定	207			
	ケーススタディ 身長の効果と比較する	214			
結 論	217	文献案内	219	キーワード	220
コーナー	220	演習問題	225	プログラミング・	

第6章 ダミー変数 ————— 233

——思った以上に便利な変数

6.1	二変量 OLS による平均値の差の評価	235			
	ケーススタディ 男女の身長差	241			
6.2	多変量 OLS におけるダミー独立変数	246			
6.3	カテゴリー変数を複数のダミー変数に変換する	250			
	ケーススタディ 国はいつ個人資産に課税するのか	254			
6.4	交互作用変数	261			
	ケーススタディ エネルギー効率	265			
結 論	271	文献案内	272	キーワード	272
コーナー	272	演習問題	274	プログラミング・	

第7章 モデルを特定する ————— 283

7.1	二次式モデルと多項式モデル	285			
	ケーススタディ 気候変動	292			
7.2	対数変換した変数	295			
7.3	処置後変数	302			
7.4	モデルの特定化	315			
結 論	321	文献案内	322	キーワード	323
コーナー	323	演習問題	324	プログラミング・	

第 II 部 現代計量社会科学の分析ツール

第 8 章	内生性と戦うための固定効果モデル, そして差の差モデル	333
8.1	データをプールすることで生じる問題	334
8.2	固定効果モデル	341
8.3	固定効果モデルを使ってみる	350
8.4	二元配置固定効果モデル	354
	ケーススタディ 貿易と同盟	358
8.5	差の差モデル	360
結 論	369	文献案内 370
		キーワード 371
コーナー	371	プログラミング・ 演習問題 375
著者・訳者紹介	385	

序 文

学生みなさんへ

——統計学を学ぶ皆さんはこの本から何を学べるのか

「つまらない昔ながらの教科書に比べれば、『社会科学のための統計分析入門』は少しはおもしろい。」——学生 A. H.

「自分の大学と大学院での経験からいわせてもらおうと、こういう教科書は本当にありがたいです。方程式とか事実とかをやたら理解させようとする教科書が多いですが、『社会科学のための統計分析入門』は、学生と教師の双方向型学習ができていて感じがしますね。」——学生 S. A.

「統計学を学び始めたときにこの本があればよかったのに。『社会科学のための統計分析入門』を最初から使っていたら大幅に時間を節約できただろうし、理解に苦しむことも少なかっただろう。」——学生 J. H.

「『社会科学のための統計分析入門』は理解しやすいうえに、忘れにくい。」——学生 M. H.

この本では、次のような重要な疑問に答えようとするときに必要とされる統計的データ分析のツールを教える。貧困対策は機能しているのだろうか。失業はインフレに影響するのだろうか。選挙での候補者の資金支出は選挙結果を左右するのだろうか。これらの疑問は、単に興味深いというだけではない。人々のため、国家のため、そして世界のためになる政策を正しいものと主張したければ、こうした疑問に正確に答えていくことが重要なのだ。

統計学を使って以上のような疑問に答えようとするなら、絶対に忘れてはいけないものがある。相関関係は因果関係と同じではないということだ。変数

X が上昇したときに変数 Y も上昇したとしよう。この場合でも、変数 X こそ変数 Y が上昇する原因であったとはいえないのだ。では、どうすれば変数 X の変化が変数 Y の変化に関係していると確証をもっていえるだろうか。この問題を明らかにするのが、この本の最重要目標だ。

この本では因果関係を解明するために必要な、3つのポイントに注目して議論を進めていく。この3つのポイントに関しては、この本以外の教科書でも目にすることがあると思う。第1に、この本では、社会科学を研究する人が最もよく使う研究ツールを取り上げている。それらの実践的な統計的テクニックを使えば、変数 X が変数 Y の原因になっていると根拠をもって主張することができる。それらのツールを使えば、みんなから一目おかれるような分析が可能になる。それらのツールを使えば、データを存分に活用することができる一方で、データから何がいて何がいてないか、どれくらい確証を持っているのかなど、データに関する限界も理解できるようになる。

この本では「現実世界の統計学」を学ぶ。現実世界の統計学では、特定の状況でときどきしか出てこないマイナーな統計的ツールはとりあえず無視する。教科書や先生がいろんなことを教えようとしすぎると、統計学はとたんにわかりにくくなってしまう。教える側が勉強して覚えた内容を何から何まで網羅しなくても、学生に統計的因果推論の全体像をしっかりと理解させることは可能だ。この本はそう主張する。

第2に、この本は、全体を通じて1つの分析枠組みに支えられている。この本では新しい概念が出てくるたび、すべてを一から学び直すような形式はとらない。核になるモデルを1つ示し、それを中心に考えていく。すなわち、1つの方程式と共通の仮定を本全体を通して扱い、調べ、探り、発展させていく。このやり方だと、本の前の部分に戻ったり、読み返したりする必要がなくなり、勉強の手間が省ける。どんなスキルを身につけるときの、あるテクニックを一回見ただけで完璧に理解することは無理だろう。そのテクニックについて、とにかく知ることだ。そのテクニックを使ってみることだ。そうすることで、きっと勝手がわかるようになる。何がどうつながっているのか、きっと理解できるようになる。そして、本当にわかったと思える瞬間が訪れる。学ぼうとするテクニックが縄跳びでも、タイピングでも、バスケのシュートでも、そしてデータ分析でも、うまくできるようになるまで何度も繰り返す必要がある。この本の全体を支える分析枠組みに食らいついていけば、すでに学んだことが、よりハッキリ理解できるようになるチャンスが増えるだろう。この本の

中で、重要だと思うことは、あえて何度でも繰り返し述べる。著者自身、繰り返すことを恐れていないからだ。

第3に、この本では政策や政治、経済に関する数多くの事例が出てくる。読者が「二段階 OLS」とか、「最尤推定」とかに全く関心がなかったとしよう。だが、「二段階 OLS」や「最尤推定」といったテクニックを理解すれば、教育政策、貿易政策、選挙結果、あるいはその他の自分が興味を持っている事柄について、自分の理解がそれまでと違ってることが実感できるはずだ。この本で扱われている例やケーススタディをみれば、この本で紹介される分析ツールが研究者によって実際に活用されていることは明白である。本書で紹介する分析ツールを使うことで、実際に優れた実証研究が生み出され続けているのだ。

『社会科学のための統計分析入門』は統計学の基礎クラスとか、計量的アプローチを教える社会科学の基礎クラスで使うことを念頭において書かれたものだ。あるいは、上級者向けの方法論クラスで、統計的データ分析のコツや背景をよりよく理解するための補助教材として使ってもらってもいい。『社会科学のための統計分析入門』には、統計学を応用して分析をするときに必要なコツやツールがすべて書かれているから、統計的データ分析の授業以外でも使える。統計学は、いまやわれわれの人生のありとあらゆるところに浸透している。統計学はエンタメにだって浸透しているし、スポーツにすら浸透してきている（アイスホッケーのプロ選手のデータを回帰分析で分析した論文なんてものを見ても、驚いてコーヒーを口から吹き出すなんてことは、最近ではなくなった）。実際に使える統計学の知識があれば、それが世界をよりよく理解する手助けになってくれるのだ。

| この本には何が書かれているのか

この本を読みこなすのに必要な予備知識はそんな多くない。各項目をていねいに説明するために、基礎的な代数をちょっと使う。微積分は必要ないが、必要があれば言及することもある。微積分を必要とする概念を使う授業の教材として、この本を使うことも可能だ。だけど、微積分を理解できていなくても、この本は完全に理解できる。

この本の最初の二章はイントロになっている。第1章では、統計学における最も重要な問いについて述べる。それは、変数間に因果関係があるということ、確率論的に、かつ正確にどうやっていえるのか、というものだ。そ

の問いに答えるのに理想的な方法は実験をすることだが、現実には実験を行うのは難しいことが多いし、われわれが知りたい疑問に実験では直接答えられないことも多い。第1章ではこの本の議論の「全体像」を示す。議論の全体像は、第1章以降の章で扱う、より詳細な議論と同様、重要な意味を持つ。

第2章では、優れたデータ分析とはどのようなものなのか、実際に分析するための基礎を提示したい。どんな統計的データ分析でも、データ分析にはソフトウェアが必要になる。これを慎重にやらないと、分析が全部ダメになってしまう。第2章では分析を記録し、データについて理解するための優れた方法とはどんなものか、解説する。

第I部の5つの章は、最小二乗法 (Ordinary Least Squares: OLS) の導入になっている。OLS は回帰分析 (Regression Analysis) ともいう。実質的には、すべての統計的データ分析は OLS か、OLS から発展してできた分析ツールによって行えるはずだ。第3章では最も基礎的な回帰分析のモデル、すなわち二変数による OLS モデルを扱う。第4章では、仮説を検証するのに OLS をどう利用するのかについて説明する。第5章から第7章までは、多変数による OLS モデルとその応用について触れる。第I部を読み終えるころには、回帰分析が理解でき、測定可能なあらゆる変数を統制できるようになっているだろう。データに曲線を当てはめたり、特定の変数の効果量がグループごとに異なるのかどうかを検証できるようにもなっているはずだ。ここまでくれば、あなたの友人も、あなたのデータ分析の腕前に感心するに違いない。

第II部では、最新の統計分析で使われている分析ツールを紹介する。これらのツールは論文を学術誌に出版するときとか、カネを稼ぎたいときに有用だ。これらのツールは多変量の OLS を発展させたものなのだが、二変数間の因果関係を識別するのにより役に立つ。第8章では、直接測定できない多くの要因をうまく扱うための、単純だが強力な手法を紹介する。第9章では、操作変数法を扱う。この手法は独立変数に影響を与えるけれど従属変数に影響を与えない、そういう変数を見つけたときに有効である。この操作変数という手法はちょっとクセが強い。だが、因果効果を浮かび上がらせるうえで、すごく有益な手法になりうる。第10章で扱うのはランダム化実験 (Randomized Experiment) だ。理論的にはランダム化実験が行えるのが理想だが、実際に実験をやってみると、注意しないといけない難問がいくつも浮かび上がってくる。第11章では、回帰不連続 (Regression Discontinuity) というツールを扱う。この手法は、一定の条件を満たした変数について、その効果を見る際に使

える。たとえば、アメリカ合衆国ではメディケアという医療費補助制度が65歳以上になると利用可能になる。そして、私立学校の中には、テストが一定の点数を超えていないと入学できない場合がある。こうした、閾値を超えたときにのみ起きる政策・決定を手がかりにして、より信頼のおける統計的分析を行うことが可能なのだ。

第III部には、章が1つしかない(第12章だ)。ここでは二値変数を従属変数にするモデルを扱っている。このモデルだと、われわれが知りたい結果は2つの値しかとらないことになる。実例をあげると、高校を卒業するかどうか(卒業できる人と、できない人の二値をとる)、失業しているかどうか(職についているか、いないかの二値をとる)、同盟を組むかどうか(2つの国は同盟条約を批准するか、しないかの二値をとる)などがある。第12章では、どうやってこういう二値のモデルにOLSを当てはめるのか、検討する。そのうえで、独立変数が二値変数であるモデルを使う際に生じる不足を補える、精巧なモデルを紹介する。

第IV部では、いろいろ有益な手法を追加的に紹介し、それまでの内容を補足する。第13章では時系列データを扱う。この章の前半では、OLSから派生したモデルについて説明し、後半では、OLSとは決定的に異なる動的モデルを紹介する。第14章では、OLSの発展形について述べ、いくつかの具体的な問題に関して議論する。第15章では、パネルデータについて詳細に扱う。そこで話の中心になるのは、パネルデータに含まれるさまざまな要素をどうやって1つにまとめるのか、という点だ。

第16章は、この本の結論である。ここでは、統計的現実主義者になるための心がけについて述べる。実際のところ、統計分析はどれくらい有益なのか、統計分析の限界はどこにあるのか、大まかに理解したいなら、この章を最初に読むといい。そして、ほかの章で扱われている概念を理解できるようになってから、もう一度この章を読み直してほしい。

| この本をどう使うか

『社会科学のための統計分析入門』は、統計的データ分析を習得してもらうために書かれたものだ。各節は「これだけは覚えよう」コーナーで終了するようになっていて、その節で学んだ重要なポイントを強調してある。「これだけは覚えよう」コーナーの内容を覚えていくと、統計学の基礎がかなり理解できるようになるだろう。重要事項は本文中、最初に登場したところでゴチックに

してある。本文の下の方で、その重要事項について簡単に解説する。これらの重要事項については、本の最後に掲載されている「用語解説」でも再度説明する。

節によっては、「理解の確認」と「議論してみよう」がついているところもある。これは取り組んだほうがいい。課題に回答することで、自分が本当に内容を理解できているのかどうか確かめられる。苦勞して統計学の勉強に取り組んだとしても、認知心理学者のいう「説明深度の錯覚 (illusion of explanatory depth)」という状態に陥ることもある。われわれは自分が思っているほどには、自分がやっていることを理解できていない、というのが「説明深度の錯覚」だ。「理解の確認」と「議論してみよう」に回答してみると、自分がどの程度理解できているのかわかるだろう。「理解の確認」は具体的なことを聞く問題で、個別に解答が存在する。解答は本の終わりに掲載している。「議論してみよう」は自由回答形式になっていて、自分が関心を持っている物事にどうやって勉強した概念を適用していくか、突き詰めて考えていくためのものだ。こうやって問題を解いていけば、統計学を学ぶために統計学をやる、みたいな状態を脱することができる。むしろ、重要な争点・事柄をよりよく理解するために統計学を使っているという状態になる。

何かを勉強するというのは、ひとりで問題に答えることだけではないことも覚えておいてほしい。先生でも、クラスメートでも、バスで隣の席に座っているイケメンでも、そういう人と一緒になって自分なりの疑問を思いつくことも、勉強していくうえでは重要だ。そうやっていくと、自分は何がわかっていないかを理解できるようになるし、いろいろな人と意見を交換するキッカケになるだろう。バスのイケメンとはそれを機に仲良くなれるかもしれない。最悪、その人はバスを降りて逃げてしまうかもしれないが……。

最後に、この本の著者はいろいろなことにこだわってほしい、この本はずいぶんくだけた言葉で書かれているな、と思うかもしれない。たしかに、この本はありきたりの統計学の本とは雰囲気が違う。でも、くだけた言葉で書かれているといっても、内容までダメだというわけじゃない。この本では、本当に優れた統計的データ分析、本当に優れた方程式、本当に優れた研究の例が出てくる。ほかの統計的データ分析の本ではお目にかからない、すごい知識も得られるかもしれない。こうすることで、この本の内容はよりとっつきやすくなり、統計学との正しい向き合い方をわかりやすく伝えられる。統計学は数理的な方程式の羅列がすべてじゃない。そうではなくて、何かに関心を持っている

人たちが現実世界から何かを学べるような、実際に使える分析ツールを提供していくのが統計学という学問なのだ。しかし、分析ツールに振り回されるようでもダメだ。この本は『社会科学のための統計分析入門』であって、『サルでもわかる統計学』じゃない。この本の内容を勉強すれば、自分や社会にとって重要な問いに答えるために、統計学をどんどん使えるようになっていくだろう。

教師の皆さんへ——統計学を学生にどのように教えればいいのか

統計学を教える教員は、学生に高望みしがちだ。学生には、統計学を使って、経済学や政策に関する重要な問いに答えられるようになってほしい。ときには、学生が信じられないような洞察をもたらして、われわれ教師を満足させてくれることもあるだろう。天の雲の切れ目から日光が射し込み、天使が歌を歌うように。毎日こういうことが起これば、どんなにすばらしいか。だが、悲しいことに、こういう経験よりも、教師がよく経験するのは、学生が混乱という立ちで苦虫を噛みつぶしたような顔をしている光景のほうだ。教室に雲が立ち込め、雨が降る、といったような。

統計学を教えると、必ずこうなるわけではない。最も重要な概念に絞って教えていけば、学生が多くのに気がつくようにヒントを与えていくことが可能だし、学生の苦虫を噛みつぶしたような顔を見ることも少なくなるだろう。不幸なことに、ありきたりの統計学や統計学の教科書は、内容が単純すぎるわりに、わかりにくいことが多かった。多くの教科書はあまりにも単純で、過去の遺物みたいな初歩的な OLS 以上のことはほとんど学ばずに、大学の一学期が終わってしまうような内容になっている。学生に統計学がどれほど使えるものか、おもしろいものかを示すこともしないで、OLS にたどり着く前に確率論に深入りして、内容が非常にわかりにくくなってしまっている教科書も多い。

『社会科学のための統計分析入門』は、われわれ教師が実際に使っている分析ツールを教えるようにすれば、最も効率的に学生に統計学を教えることができるはずだ、という信念に基づいて書かれた本である。われわれ教師は、回帰分析を中心とした、とくに近年は実験と因果推論に焦点を絞った、分析ツールを使っている。学生がこういう基礎的なコンセプトを理解できるようになれば、彼ら・彼女らも正式に統計的データ分析に関する真剣な議論に参加す